Journal of Experimental Psychology

ARTHUR W. MELTON, Editor
AIR FORCE PERSONNEL AND TRAINING RESEARCH CENTER
LACKLAND AIR FORCE BASE
SAN ANTONIO, TEXAS

CONSULTING EDITORS

Judson S. Brown, *State University of Iowa*
Cletus J. Burke, *Indiana University*
Robert M. Gagné, *AF Personnel and Training Research Center*
W. R. Garner, *The Johns Hopkins University*
Frank A. Geldard, *University of Virginia*
James J. Gibson, *Cornell University*
Clarence H. Graham, *Columbia University*
David A. Grant, *University of Wisconsin*
Lloyd G. Humphreys, *AF Personnel and Training Research Center*
Arthur L. Irion, *Tulane University*
Donald B. Lindsley, *University of California, Los Angeles*
Neal E. Miller, *Yale University*
Kenneth W. Spence, *State University of Iowa*
Benton J. Underwood, *Northwestern University*
Delos D. Wickens, *Ohio State University*

Lorraine Bouthilet, Managing Editor

CONTENTS

Stimulus Correlates of Visual Pattern Recognition: A Probability Approach: P. M. FITTS, M. WEINSTEIN, M. RAPPAPORT, N. ANDERSON, AND J. A. LEONARD 1

Relationship Between Recognition Accuracy and Order of Reporting Stimulus Dimensions: D. H. LAWRENCE AND D. L. LABERGE 12

Color Coding in a Visual Search Task: B. F. GREEN AND L. K. ANDERSON 19

Time Continuously on Target as a Function of Distribution of Practice: L. E. BOURNE, JR. AND E. J. ARCHER 25

Stimulus Sequence and Concept Learning: M. H. DETAMBEL AND L. M. STOLUROW 34

Probabilities of Success and Amounts of Work in a Multichoice Situation: M. H. DETAMBEL 41

Acquisition, Extinction, and Spontaneous Recovery as a Function of Percentage of Reinforcement and Intertrial Intervals: D. J. LEWIS 45

Interrelationships of Successive and Simultaneous Discrimination: A. J. NORTH AND M. JEEVES 54

Acquisition of a Conditioned Response as a Function of Forward Temporal Contiguity: M. E. FITZWATER AND R. S. THRUSH 59

Effects of Magnitude of Temporal Barriers, Type of Goal, and Perception of Self: C. M. SOLLEY AND R. STAGNER 62

The Estimation of Loudness by Unpracticed Observers: S. S. STEVENS AND E. C. POULTON 71

Discriminability and Stimulus Generalization: N. GUTTMAN AND H. I. KALISH 79

American Psychological Association

Vol. 51 No. 1 January 1956

JOURNAL OF EXPERIMENTAL PSYCHOLOGY

The JOURNAL OF EXPERIMENTAL PSYCHOLOGY is published monthly, two volumes per year, by the American Psychological Association, Inc., at Prince and Lemon Sts., Lancaster, Pa. The subscription rate per volume is $8.00, or $16.00 annually. Single copies are $1.50. Subscriptions, orders, address changes, and business communications should be addressed to the American Psychological Association, Inc., 1333 Sixteenth St. N.W., Washington 6, D. C.

This JOURNAL publishes original experimental investigations which are intended to contribute toward the development of psychology as an experimental science. Studies with normal human subjects are favored over studies involving abnormal or animal subjects. Studies in applied experimental psychology or engineering psychology may be accepted if they have broad implications for experimental and theoretical psychology. Normally, articles of a length exceeding 15 printed pages cannot be accepted. However, an integrated series of studies accomplished simultaneously (e.g., most doctoral dissertations) must be presented in a single article, rather than in a series of articles.

Articles are published in the order of their receipt, except in rare circumstances. Authors are supplied with 50 free offprints without covers. All of the cost of an author's alterations in galley proof is charged to the author. Priority in publication is given to articles whose authors assume the full cost of publication. In 1956 the publication cost is $20 per printed page. Authors of priority publications receive no free offprints.

Address all articles submitted for regular or priority publication to the editor, Arthur W. Melton, Air Force Personnel and Training Research Center, Box 1557, Lackland Air Force Base, San Antonio, Texas. All manuscripts should be submitted in duplicate. Original figures are prepared for publication; duplicate figures may be photographic or pencil-drawn copies. Authors are cautioned to retain a copy of the manuscript to guard against loss in the mail.

To save printing costs supplementary material to articles in this Journal are deposited with the American Documentation Institute. Copies may be obtained by ordering the Document Number given in the footnote reference to ADI. Orders should be sent to ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress, Washington 25, D. C. Orders *must* be accompanied by checks or money orders payable to the Chief, Photoduplication Service, Library of Congress.

Manuscripts submitted to the JOURNAL *must* adhere to the conventions concerning reference citations, preparation of tables and figures, manuscript format, etc. as described in the Publication Manual of the American Psychological Association (*Psychol. Bull.*, 1952, 49 [4, Suppl.], 389–449). In particular, the organization of the manuscript should be text, references, footnotes, tables (one to a page), figure titles (typed on separate page), and figures. All typed material *must* be double-spaced with ample margins. The term "subject," with its various forms, is abbreviated *S*, *Ss*, *S*'s, and *Ss*'; analogous abbreviations are used for "experimenter" (*E*) and "observer" (*O*). When in doubt about the practices of the JOURNAL, authors should examine a recent issue.

Entered as second-class matter, February 6, 1937, at the post office at Lancaster, Pa., under the Act of March 3, 1879. Acceptance for mailing at the special rate of postage provided for in paragraph (d-2), Section 34.40, P. L. & R. of 1948, authorized September 11, 1947.

Send address changes to 1333 Sixteenth St. N.W., Washington 6, D. C. Address changes must reach the Subscription Office by the 15th of the month to take effect the following month. Undelivered copies resulting from address changes will not be replaced; subscribers should notify the post office that they will guarantee second-class forwarding postage. Other claims for undelivered copies must be made within four months of publication.

Copyright © 1956 by the American Psychological Association, Inc.

Journal of
Experimental Psychology

VOL. 51, NO. 1 JANUARY, 1956

STIMULUS CORRELATES OF VISUAL PATTERN RECOGNITION: A PROBABILITY APPROACH [1]

PAUL M. FITTS, MEYER WEINSTEIN,[2] MAURICE RAPPAPORT,[3] NANCY ANDERSON, AND J. ALFRED LEONARD [4]

The Ohio State University

In studies of visual pattern perception it has been customary to specify stimuli in terms of their physical characteristics. Among the stimulus characteristics that have been singled out for special study are regularity and symmetry (**14**, **19**), area, perimeter, and maximum dimensions (**6**), number of elements (**8**), number of angles and line segments (**12**), number of indefinitely extended straight lines (**4**), and texture (**9**).

It is also possible to classify visual stimuli in probability or statistical terms by reference to a parent population and to the sampling process by means of which a set of figures is selected from this population. A probability approach to the study of perceptual processes is suggested by recent developments in information theory and in statistical learning theory, and has been specifically proposed by Attneave (**2**, **3**). This approach represents an extension of existing methods and theories. It offers promise as a supplement to methods and theories that emphasize habits and expectancies, dynamic field effects, or the role of specific stimulus features in perception.

Studies of the effects of language habits (**16**, Chap. 10) and of the influence of context and verbal instructions (**5**) on visual perception are in a sense concerned with factors that influence the probabilities of occurrence of responses to visual stimuli. The present approach makes available to those who wish to pursue this line of investigation a method for precise quantitative control over the probability characteristics of the stimuli, plus a method for generating an almost unlimited number of stimuli to which there are few stereotyped or pre-established responses.

A probability approach to the study of perceptual processes also makes

[1] This research was conducted in the Laboratory of Aviation Psychology of The Ohio State University through support from a contract between the Office of Naval Research and the OSU Research Foundation (Contract No. Nonr-495(02). The authors are listed serially in the order of their original association with the project. Dr. Oscar S. Adams also contributed to the early formulation of the concepts and methods underlying this research.

[2] Now at the Aero Medical Laboratory, Wright-Patterson Air Force Base, Dayton, Ohio.

[3] Now at the Stanford Research Institute, California.

[4] Visiting Research Associate from the Medical Research Council, Applied Psychology Research Unit, Cambridge, England.

available new methods for the study of relations among the parts of visual patterns and for the study of visual field effects. It is equally well adapted to the quantitative specification of temporal sequence effects and to the specification of the spatial dependencies among the parts of figures.

The approach outlined in the present paper also provides a supplemental method that can be used by investigators who are interested in examining the characteristics of individual stimulus patterns and in studying the formation of associations between specific stimuli and specific responses. In other words, *O*'s response to a particular visual pattern may be considered to be a function both of the unique characteristics of that pattern and of the probabilities associated with the class of figures of which it is a member. The conventional approach emphasizes differences within classes of figures, whereas the probability approach emphasizes differences between classes. As an illustration, in the present paper within-class variations among visual patterns are treated as random variations around sample means, but they could also be studied in their own right. The use of special sampling procedures that have the effect of restricting the set of stimuli to a homogeneous subclass of visual patterns represents an intermediate step in which specific types of differences among figures can be investigated.

Since a probability approach to perception is of general interest to students of learning as well as to those interested in perception, we shall describe the method employed in generating statistically definable stimuli and discuss briefly the concepts employed in specifying stimuli in probability terms before turning to the results of two specific experiments employing these methods. The method employed in measuring pattern recognition, and the empirical and rational bases for the selection of a recognition task for the study of perceptual processes, have been presented in the preceding article in this series (1).

Statistical Properties of Visual Figures

Procedures for generating stimuli.—One of the most distinctive properties of any visual figure is its contour or outline. In order to deal quantitatively with this property of visual forms, it is convenient to quantize the contour, i.e., to permit it to change in discrete steps rather than continuously. Quantization can be achieved by constructing solid figures in a two-dimensional matrix, the cells of which constitute the unit variations in figure contour. As an illustration, consider two symmetrical matrices, 4×4 and 8×8, in which the individual cells of the 64-cell (8×8) matrix are one-fourth the area of those of the 16-cell matrix (see Fig. 1). Solid figures are generated within these matrices as follows: Starting with the left-hand column or at any other convenient position, a number greater than zero, but not exceeding the number of possible quantizing steps in that row or column, is determined by application of some sampling

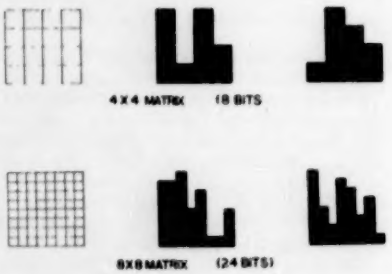

Fig. 1. Construction of metric figures. Two matrices, consisting of 4×4 and 8×8 cell units, are shown with illustrations of asymmetrical figures constructed by sampling contour details with replacement (see the two center patterns) and by sampling details without replacement (right-hand patterns).

procedure. This number determines the magnitude of area contained within that portion of the figure. The sampling procedure is continued until a magnitude has been determined for each portion of the figure. Corresponding black figures such as those shown in the first column of Fig. 1 are then constructed on a white background. Solid figures constructed in this manner have a maximum of 4 and 8 contour details, respectively. In order to emphasize the fact that visual patterns constructed in this manner, both individually and collectively, are capable of precise quantitative specification, we propose to call them *metric figures*,[5] where metric is used in the sense "pertaining to measurement."

If each detail of the contour of a solid metric figure of the sort described above is determined independently by a random sampling procedure, it is possible to generate $N = r^c$ different figures in an $r \times c$ matrix—256 in the 4×4 matrix, and 16,777,216 in the 8×8 matrix. Adopting the convenient information notation, the random selection of one figure from among each of these two populations can be said to generate $\log_2 N$, or 8 and 24, bits of information, respectively. Considered from an alternative viewpoint, the selection of the magnitude of any one contour detail (column height) in the 4×4 matrix involves the choice of one in four equally-likely alternatives, or 2 bits of information, and thus the independent selection of four columns again is seen to require 4×2 bits, or 8 bits of information.

The average amount of information required to specify one from among all the possible solid figures that can be generated in a given matrix by a random sampling process is proposed as a quantitative (statistical) measure of the degree of *complexity* characterizing the figures generated in that matrix. From this viewpoint all solid figures generated in any matrix of a particular size are defined as equally complex.[6] Complexity, as here defined, is a statistical property of the population from which stimuli are selected by E; the definition is independent of S's performance, independent of the sampling operation, and independent of the particular stimulus patterns that result from the sampling operation. Other statistical characteristics of stimuli such as the ratio of actual to potential information, as well as certain physical features, will also need to be specified in a complete description of a set of stimuli.

Sampling rules.—The maximum number of different solid metric figures that can be generated in any matrix is that number resulting from the use of a random sampling procedure. Any nonrandom sampling procedure will produce fewer total figures, and by definition, figures possessing some degree of relative *redundancy*. The degree of relative redundancy characterizing a particular class of metric figures is here specified in the conventional way as one minus the ratio of actual to possible information per figure, and is a function of the degree of restriction characterizing the sampling process.

Many different nonrandom sampling procedures can be employed in generating metric figures and each has associated with it a particular degree of redundancy. As one illustration, if the constraint is such that each contour detail appears once and only once in a given figure, then the total set of figures that can be generated in an $n \times n$ matrix is reduced from n^n to $n!$ and the relative redundancy of the resulting figures is one minus the ratio of $\log n!$ to $\log n^n$. In the case of a 4×4 matrix this amounts to the introduction of 42.75% redundancy. This particular sampling constraint, which corresponds to sampling of contour details without replacement, insures that all figures generated in any $n \times n$ matrix will have equal areas and an equal number of corners. An equal-area restriction can also be imposed in other ways. As an illustration, the 256 solid figures that can be generated in a 4×4 matrix by a random process vary in total area from 4 to 16 units. If only figures of 10 units

[5] The term *histoform* has also been suggested as an appropriate name.

[6] Alternatively stimulus complexity could be defined in terms of actual rather than maximum possible information; however, the definition recommended here permits a sharper distinction to be made between complexity and redundancy.

of area (the mean area) are used, thed only 44 different figures are possible ann relative redundancy becomes 31%. If figures are restricted to any one of the other 12 possible areas then redundancy will be greater than 31%.

Redundancy can also be introduced by making one part of the contour of the figure partially or completely dependent on some other part. As an illustration, a figure can be constructed in one-half of a matrix and either (*a*) a simple reproduction, or (*b*) a mirror-image reproduction can be formed in the other half (see Rows B, C, E, and F in Fig. 3). In this example both types of dual-contour metric figures have the same relative redundancy (50%). Redundancy also results when different contour details are sampled with unequal frequency.

From this brief discussion it should be clear that it is possible to study not merely the effects of redundancy per se, but also the effects of introducing redundancy by different methods. There is no a priori reason for believing that all types of redundancy necessarily have a similar effect on perceptual processes, and the problem can be treated as involving the effects of different methods of stimulus coding.

Addition of noise.—The concepts of noise level, and of signal-to-noise ratio, have been widely used in auditory research and in the specification of communication systems. A signal is customarily defined as conveying some kind of message, and a noise as any source not coherent with this message. The techniques described above for generating metric figures lend themselves readily to the precise specification of different kinds of visual perturbation or noise effects. As an illustration, a metric figure can be distorted by the addition or deletion of unit increments to its contour (see Fig. 2A), the amount of perturbation and the particular details which are to be modified again being determined by specified laws of probability. Another type of perturbation is one affecting all units of the figure and of its background, for example, a noise source which would change some of the black cells of the figure into white cells and some of the white cells of the background into black cells (see Fig. 2B).

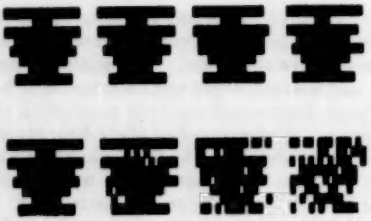

FIG. 2. Two types of visual noise. The first figure in Rows *A* and *B* is undistorted or noise-free. In the remaining figures of Row *A* three levels of visual noise (6.25%, 12.5%, and 25%) have perturbed the cells adjacent to the contour lines; in the remaining figures of Row *B* three comparable noise levels have affected all cells of the matrix. Noise level is here defined as the probability that the brightness of a given cell will be reversed.

It is possible to study the effectss of different types as well as different levels of noise. A topic of special interest to engineering psychology is the possible use of different types of redundancy to combat the effects of various kinds of noise. At a theoretical level the systematic study of the effects of noise offers a new approach to such problems as form constancy and the perception of designs masked by more complex designs (10).

Specific physical features of stimuli.—In constructing metric figures by means of the statistical procedures described above it is necessary to specify certain physical features of all the figures in a sample. Among these are (*a*) over-all area, (*b*) the ratio of maximum height to width, (*c*) the dimensions of the cells of the matrix, (*d*) the number of quantizing units assigned to each of the dimensions of the matrix, (*e*) the spatial orientation of the figures, and (*f*) brightness-contrast relations. The effects of these and other physical features can be studied concurrently with the effects of the statistical properties of the metric figures.

Sample size.—The size of the sample of figures used in a particular experiment is a central problem for the present meth-

odology. The results of several completed investigations (**17, 18**) indicate that individual differences among the figures comprising a sample are of the same order of magnitude as are individual differences among *S*s. Interactions between different samples of figures and experimental effects have generally been found to be nonsignificant. The use of single samples of eight figures has been found to be satisfactory for perceptual learning and transfer studies, and for testing hypotheses when experimental effects are relatively large. The determination of precise psychophysical functions with highiy practiced *S*s, however, appears to require the use of larger samples, or the use of experimental designs in which different randomly drawn samples of figures are used for each *S* as well as for each condition studied.

Verbal associations.—One feature of the metric figures here described, which may be of special interest for students of learning, is their low degree of meaningfulness. Unlike other commonly used visual stimuli, such as colors and geometric symbols, few of the metric figures elicit stereotyped verbal associations. In some respects they may be less meaningful than nonsense syllables. Thus, they may prove to be useful in many experimental situations where alternative sets of homogeneous stimuli are needed.

A research program.—The preceding discussion does not exhaust the ways in which the statistical properties of metric figures can be varied. However, it illustrates some of the possibilities offered by a probability approach to the study of form perception. The three concepts, complexity, redundancy, and noise, together with certain associated physical features of stimulus patterns, provide the basis for an extensive program of research. Two studies, carried out as part of such a program, are reported here.

Experiment I

This experiment was designed to investigate the effects of three stimulus factors, (*a*) single vs. double vs. bilaterally symmetrical figures, (*b*) figures constructed by sampling of contour details with replacement (randomized sample) vs. sampling without replacement (constrained sample), and (*c*) horizontally vs. vertically oriented figures. Speed in recognizing visual patterns was the dependent variable.

Method

Stimulus variations.—Three types of metric figures were employed. The first was an asymmetrical single-contour figure constructed in an 8 × 8 matrix (see Fig. 3A and 3D). The second was an asymmetrical double-contour figure constructed in a 16 × 8 cell matrix (Fig. 3B and 3E). The third was a symmetrical double-contour figure also constructed in a 16 × 8 cell matrix (Fig. 3C and 3F). All three types of figures were constructed in matrices of equal over-all area (.8 × .8 in.), by a sampling process that involved the same amount of actual information. However, since the figures with double contours could have contained twice as many unique 3-bit details as the single contour figures they are 50% redundant with respect to the single-contour figures. These three types of figures permit a comparison of the effect of two methods of introducing redundancy, both of which involve replication of a part of a contour, with amount of actual information per figure held constant.

Samples of eight metric figures of each of the three types were generated by two different sampling procedures. Under the first procedure it was specified that each contour detail of a figure be selected independently by reference to a table of random numbers with the restriction that each detail appear with equal frequency in the sample of eight figures but not necessarily in each individual figure. This procedure is

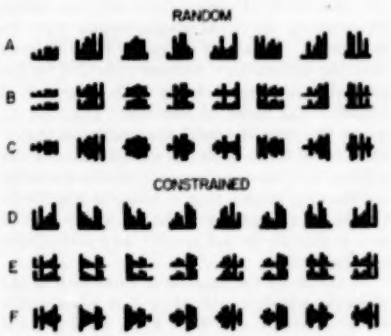

Fig. 3. The six sets of metric figures employed in Exp. I. The figures in Rows *A*, *B*, and *C* were generated by randomized sampling; those in Rows *D*, *E*, and *F* were generated by sampling without replacement.

equivalent to random sampling of contour details without replacement from a finite population of 64 details (eight of each category of height). It should be noted that the first figure selected in this manner could be any one of almost seventeen million figures. Figures generated by this procedure will be referred to hereafter as randomized metric figures. Under the second sampling procedure it was specified that each contour detail be selected by means of a random process with the important restriction that each detail appear once and only once in the contour of each figure. Approximately 40,000 figures can be generated in an 8×8 cell matrix by this procedure, which is equivalent to sampling of contour details without replacement from a finite population of eight different details. All figures generated by this process have equal area. They will be referred to hereafter as constrained metric figures. The two samples of figure contours were reproduced in each of the three ways described in the preceding paragraph. Seven replications were made of each figure.

Figures were surrounded by a .4-in. white border and were mounted in 1.6×1.6-in. plastic covers to protect them from being defaced by handling.

Figures were used in two orientations, horizontal and vertical. The horizontal orientation is that shown in Fig. 3 and will be called the 3 o'clock orientation. The vertical or 6 o'clock position involved a 90° rotation of each figure from the 3 o'clock positions. At 6 o'clock the base of the asymmetrical figure is on the left and the distinctive contour on the right.

Recognition task and apparatus.—The recognition task employed in the present study is an improved version of a sorting task employed in an earlier study (1) in which the relations among 11 measures of visual discrimination were investigated. Results of the earlier study indicate that the performance measure provided by the present task agrees closely with measures of legibility and speed of learning. The task requires the type of visual discrimination that is the basis of man's ability to identify a particular pattern from among a group of similar patterns, but does not require that he name or describe it. One of the advantages of the use of this type of task is that *S*s do not require long preliminary training periods.

A sketch of the apparatus employed in the recognition task, as viewed by *S*, is given in Fig. 4. It consisted of a 36×40-in. sorting board on which were arranged six rows of eight figures each. The front edge of the board was 40 in. from the floor and the board sloped upward at 10°. Figures were supported by thin wooden strips so that they were approximately perpendicular to the line of sight, and could easily be turned over. At the top of the board was a

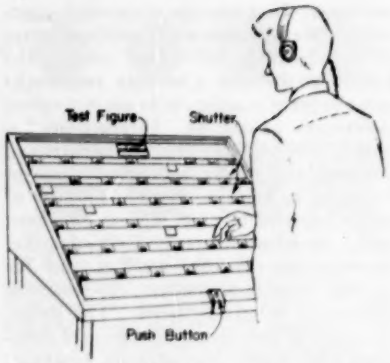

FIG. 4. Schematic drawing of the test board with the shutter open. The *S* has located and turned over the test figure in the first four rows.

holder for a test figure. At the bottom was a push-button switch.

A six-bladed shutter could be positioned so that it shielded all figures except the test figure from *S*'s view. When released by *E*, the shutter slid down into a position which exposed all figures. A .01-sec. timer was automatically started by the opening of the shutter.

Cushioned headsets were provided through which *S*s heard broad-band noise at about 60 db. Illumination was uniform and nonglare.

The sorting board was arranged so that the same sample of eight homogeneous figures appeared in each row with uniform orientation. The order of figures along each row was determined by independent shuffling, the only restriction being that the same figure not appear in the same position in adjacent rows. The *S* stood in front of the board and wore the headset. He was given 2 sec. to inspect the test figure. The shutter was then opened, and he began to search for the test figure beginning with the top row. As soon as he located the figure he flipped it over, using his right hand, and went on to the next row. He could verify the identity of the figure by means of an easily read color code on the back. After finding the sixth figure *S* pushed the button at the bottom of the board. This stopped the timer. The method of correction was employed whenever errors occurred, but errors were very infrequent. Completion of one sorting board, requiring the recognition of six replications of the test figure, constituted one item of the test.

A trial consisted of a sequence of eight items, during which each figure in the sample of eight served once as the test figure. The order of the items in a trial was determined independently for each *S* by shuffling the set of eight test figures.

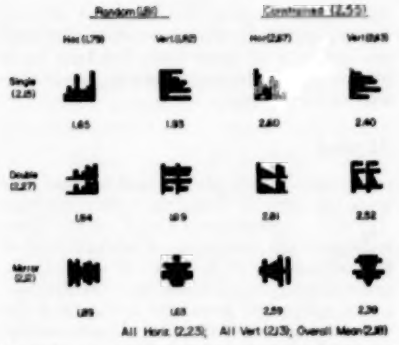

FIG. 5. Average recognition time in seconds per figure for the three factors studied in Exp. I.

After each trial the figures on each row of the sorting board were also taken up, shuffled, and replaced.

Each *S* was given six trials a day for two days. At the end of this time he had been tested once on each of the eight figures in each of the two samples, reproduced in each of the three ways, and oriented in each of the two positions. The order of the 12 conditions was determined by reference to a table of random numbers and was different for each *S*. This procedure was repeated on two additional days, so that the entire experiment was replicated twice over a total period of four days. The first replication was treated as a practice period and only the data for the last two days were analyzed.

Subjects.—Twelve volunteer male college students served as *S*s. They were paid hourly wages.

Scores.—The time required by *S* to complete each item was recorded. This measure included the time required to turn over the six figures constituting an item (manipulation time) plus the time required to stop the clock after the last figure was located. The average time to complete an item, divided by six, was taken as the time in seconds required on the average to locate and turn over a single figure. This time per individual figure is the measure reported in the results section. Two clocks were used. One was read and reset by *E* after each item; the other was allowed to accumulate time until the end of a trial. At the end of each trial *S* was told his over-all time for that sample of figures to the nearest .1 min. In early pilot studies time was recorded separately for each of the six subparts of an item. These times were found to be unimodal and nonskewed, and the present timing procedure was adopted because of its greater convenience.

Errors were so few that they were not recorded.

Results

The mean identification times per figure, pooled across samples of eight figures and across the 12 *S*s, are shown in Fig. 5 for each of the 12 experimental conditions. By far the largest difference in recognition time was that attributable to the difference between randomized and constrained figures. On the average, the recognition of a constrained figure required .7 sec. longer than did the recognition of a randomized figure. Each of the 12 *S*s had faster mean recognition times for the randomized figures than for the constrained figures, and each of the three types of figures in both orientations were recognized more quickly than were corresponding figure types in the randomized sample.

A four-variable analysis of variance, with *S*s as one variable, was carried out and evaluated as a mixed model (**15**, p. 330). The results are summarized in Table 1. The triple interaction of the three stimulus variables was significant, as was the interaction of orientation with sampling rule and the interaction of orientation with type of figure. The difference

TABLE 1

ANALYSIS OF VARIANCE FOR EXP. I

Source	*df*	Mean Square	*F*
Subjects (*S*)	11	.513	
Type of figure (F)	2	.280	7.78**
Orientation (O)	1	.386	5.59*
Sampling rule (R)	1	19.982	112.89**
S × F	22	.036	
S × O	11	.069	
S × R	11	.177	
F × O	2	.209	3.67*
F × R	2	.026	
O × R	1	.596	18.06**
S × F × O	22	.057	
S × F × R	22	.064	
S × O × R	11	.033	
F × O × R	2	.232	4.14*
S × F × O × R	22	.056	
Total	143		

* Significant at $P < .05$.
** Significant at $P < .01$.

between the two sampling procedures, which accounted for more than half of the total sum of squares, was also significant as were the other two main variables, type of figure and orientation.

In order to examine further the effects underlying the triple interaction, and the interaction of orientation with each of the other main variables, the data for each of the two sampling methods, random and constrained, were examined by means of separate analyses of variance.

For the randomized figures the interaction of orientation and type of figure was highly significant ($P < .01$). Referring to Fig. 5, it will be seen that single-contour randomized figures were identified about .3 sec. more rapidly when they were oriented horizontally or base down, the mirror-image figures were identified more than .2 sec. faster when they were oriented vertically, and the double figures were relatively unaffected by orientation.

For the constrained figures, vertical orientation was superior by .2 sec. or more in the case of all three types of figures and this effect was highly significant ($P < .01$). None of the interactions among the constrained figures were significant.

It should be noted that the main difference in scores for the two sampling procedures was the effect of orientation on the single-contour figures. Figures of this type belonging to the randomized sample were identified more quickly on the average when viewed base down, whereas corresponding figures from the constrained sample were recognized more rapidly when viewed with the base to the left.

Experiment II

In Exp. II the investigation of the effect of orientation was extended to include figures oriented in each of the four cardinal positions. Randomized sampling was employed with all figures since the significant interaction of orientation and type of figure which had been found previously with the random sample appeared to warrant further study.

Method

The same sample of eight random figure contours, the same 12 *Ss*, and the same recognition task were employed as in Exp. I. Each *S* was tested with the three types of figures in each of four orientations (3, 6, 9, and 12 o'clock), i.e., under 12 experimental conditions. One familiarization session was given with the figures in the two new orientations (9 and 12 o'clock) in order to provide the same amount of practice on these new orientations as had been given previously on the 3 and 6 o'clock orientations. The familiarization session was followed by two test sessions. Six experimental conditions were employed per session. The order of the 12 conditions again was random and different for each *S*.

Results

Mean recognition times within samples pooled across *Ss* and figures are given in Table 2. The results of an analysis of variance of the data are summarized in Table 3.

TABLE 2

Mean Recognition Time per Figure in Seconds for the Conditions of Exp. II

(Data pooled for 12 *Ss* and 8 figures per condition)

Type Figure	Orientation*			
	3 O'clock (1.55)	6 O'clock (1.44)	9 O'clock (1.52)	12 O'clock (1.45)
Single (1.51)	1.57	1.47	1.50	1.50
Double (1.57)	1.61	1.52	1.60	1.56
Mirror (1.39)	1.46	1.34	1.44	1.30
	Over-all Mean (1.49)			

* The 3 o'clock orientation is that shown in Fig. 3; other orientations represent successive 90° clockwise rotations. The 3 and 9 o'clock positions have base line horizontal, 6 and 12 o'clock position have base line vertical.

TABLE 3

ANALYSIS OF VARIANCE FOR RANDOM FIGURES IN EXP. II

(Same 8 figure contours under each condition)

Source	*df*	Mean Square	*F*
Subjects (*S*)	11	.316	
Type of figure (F)	2	.425	17.71**
Orientation (O)	3	.089	4.24*
$S \times F$	22	.024	
$S \times O$	33	.021	
$F \times O$	6	.013	
$S \times F \times O$	66	.024	
Total	143		

* Significant at $P < .05$.
** Significant at $P < .001$.

There was continued improvement in the speed with which *S*s recognized the figures in these particular samples. The average improvement over comparable conditions in Exp. I represented a 17% reduction in time, and for individual conditions ranged from 5 to 24%.

Orientation and type of figure both led to significant variations in recognition time. The mirror-image figures were recognized most rapidly, and the double-contour, asymmetrical figures were recognized most slowly. The two vertical orientations (major axis at 6 or 12 o'clock) were sorted significantly ($P < .05$) faster than were the two horizontal orientations. However, the differences between the two vertical (6 and 12 o'clock) and between the two horizontal (3 and 9 o'clock) orientations of the figures were negligible in magnitude.

DISCUSSION

The use of a randomized vs. a constrained sampling procedure for generating figures resulted in the largest effect on recognition time of any of the stimulus variables included in these two experiments. On the basis of the relation of response time to stimulus uncertainty (**11, 13**) it might have been predicted that the constrained figures would be recognized more quickly, since the amount of information per constrained figure is much less than that for the random figures. Exactly the opposite effect was found. In interpreting these results we attribute the effect not to the fact that information content of the metric figures was reduced by the introduction of redundancy, but rather to the fact that an unfavorable form of redundancy was introduced in the figure contour.

At a descriptive level the constrained figures certainly give the appearance of being more similar to each other than the random ones. It has already been stated that they do not differ in area or in the number of corners; there is also the absence of certain distinctive cues such as identical adjacent details. Increased similarity nevertheless is not a wholly satisfactory answer for these results, except in the sense that it raises a number of questions. The primary question concerns the stimulus determinants of similarity. A further question is the extent to which the effects found with constrained figures result from presenting *S* with sets of figures simultaneously or from properties of the single figures. Incidental observations from other studies carried out in this program suggest that when a constrained-type figure appears by chance as a member of a random sample it may still show a long sorting time. In any case, there is reason to believe that the use of different sampling procedures may prove a useful method for a quantitative analysis of perceptual similarity.

The results for the two sampling procedures and for the three types of figures all support the general hypothesis, for the noise-free case, that the introduction of redundancy may either facilitate or hinder pattern recognition, depending on the way in which it is introduced. The problem posed by these findings is that of determining the effects of different types and degrees of redundancy and of examining the effects of complexity as variables influencing pattern recognition.

The finding regarding the superior performance with vertically oriented, bilaterally symmetrical figures supports one of the incidental findings from an earlier study (1) in this series, and might have been predicted in the light of observations from other types of investigations (14, 19). Data from tracking studies (7) also indicate that *S*s show greater facility in responding concurrently to two stimuli that are symmetrical around a vertical axis (left-right separation) than around a horizontal axis. Life-long experience in dealing with vertical bilaterally symmetrical figures may partially account for this effect. Many everyday objects such as people, trees, and milk bottles, present this type of double contour. It is also possible that the effect may be related to anatomical redundancy (the visual system itself is bilaterally symmetrical around a vertical axis) or to greater facility in responding to peripheral details in the lateral than in the vertical areas of the visual field.

A word of caution is called for at this time regarding the possible limitation of the present findings to recognition-type tasks. In a previous study (1) recognition performance with a set of visual patterns was found to correlate moderately highly with errors and trials in learning names for the same visual patterns. However, just as different ways of introducing redundancy appear to have differential effects on recognition performance, so redundancy and other stimulus variables may have different effects on other kinds of perceptual processes. Attneave's (3) results with somewhat different types of patterns and reproduction tasks indicate such differences.

A comparison of the data for the two experiments suggests that learning proceeded at differential rates for different classes of figures. For example, 11 of the 12 *S*s showed a larger percentage improvement in speed of recognition during Exp. II, relative to their own earlier performance in Exp. I, for the vertically oriented than for the horizontally oriented single figures. Similarly, 11 of the 12 *S*s showed a greater percentage gain from Exp. I to Exp. II for the bilaterally symmetrical than for the asymmetrical figures. Even stronger evidence for differential learning effects results when we pool the data for all *S*s and examine the results separately for the eight basic stimulus patterns in each sample of figures. Here we find for each of the eight figures that percentage learning changes were larger for bilaterally symmetrical reproductions than for asymmetrical reproductions of the same figure contour. These results suggest that *S*s learn to use different methods or different types of cues in identifying different types of figures and that both rate of learning and final level of proficiency in a recognition task may vary with type of cues available for identifying particular figures.

Summary

A method is described for generating classes of figures that can be specified precisely in probability terms. They are referred to here as *metric figures* in order to emphasize their quantitative nature. The use of samples of such figures permits the systematic study of stimulus variables that characterize classes of figures. This statistical approach to the study of stimulus correlates of perceptual processes is supplemental to an approach that emphasizes characteristics peculiar to individual figures; in fact, both kinds of variables can be studied in a single experiment.

Two experiments were conducted using samples of metric figures constructed in accordance with probability concepts. The perceptual task was one demanding speed in recognizing a particular pattern when it was presented as one of a set of alternative patterns.

Figures generated by a random process were found to be recognized much more rapidly than were constrained figures generated by sampling contour details without replacement. This result is interpreted as indicating a detrimental effect on recognition performance of a particular form of redundancy rather than a detrimental effect of a decrease in information per se.

At the end of a period of training the symmetrical figures and the vertically oriented figures were identified more rapidly than were single- or double-contour asymmetrical figures, or horizonally oriented figures of equal complexity. Theoretical implications of these results are discussed.

The present approach offers a contribution to the methodology of the study of visual percep-

tion and adds several new classifications to a rather small list of psychological variables that can be defined on the basis of sampling operations. It is suggested that the metric figures employed in these studies may prove to be useful in other studies where relatively homogeneous stimuli with few stereotyped verbal assocations are required.

REFERENCES

1. ADAMS, O. S., FITTS, P. M., RAPPAPORT, M., & WEINSTEIN, M. Relations among some measures of pattern discriminability. *J. exp. Psychol.*, 1954, **48**, 81–88.
2. ATTNEAVE, F. Some informational aspects of visual perception. *Psychol. Rev.*, 1954, **61**, 183–194.
3. ATTNEAVE, F. Symmetry, information, and memory for patterns. *Amer. J. Psychol.*, 1955, **68**, 209–222.
4. BIRKHOFF, D. B. *Aesthetic measure.* Cambridge: Harvard Univ. Press, 1933.
5. CARMICHAEL, L., HOGAN, H. P., & WALTER, A. A. An experimental study of the effect of language on the reproduction of visually perceived forms. *J. exp. Psychol.*, 1932, **15**, 73–86.
6. CASPERSON, R. C. The visual discrimination of geometric forms. *J. exp. Psychol.*, 1950, **40**, 668–681.
7. FITTS, P. M., & SIMON, C. W. Some relations between stimulus patterns and performance in a continuous dual-pursuit task. *J. exp. Psychol.*, 1952, **43**, 428–436.
8. FRENCH, R. S. Identification of dot patterns from memory as a function of complexity. *J. exp. Psychol.*, 1954, **47**, 22–26.
9. GIBSON, J. J. *Perception of the visual world.* New York: Houghton Mifflin, 1950.
10. HANAWALT, N. G. The effect of practice upon the perception of simple designs masked by more complex designs. *J. exp. Psychol.*, 1942, **31**, 134–148.
11. HICK, W. E. On the rate of gain of information. *Quart. J. exp. Psychol.*, 1952, **4**, 11–26.
12. HOCHBERG, J., & MCALISTER, E. A quantitative approach to figural "goodness." *J. exp. Psychol.*, 1953, **46**, 361–364.
13. HYMAN, R. Stimulus information as a determinant of reaction time. *J. exp. Psychol.*, 1953, **45**, 188–196.
14. KOFFKA, K. *Principles of gestalt psychology.* New York: Harcourt, Brace, 1935.
15. MCNEMAR, Q. *Psychological statistics.* New York: Wiley, 1955.
16. MILLER, G. A. *Language and communication.* New York: McGraw-Hill, 1951.
17. RAPPAPORT, M. The role of redundancy in discrimination of visual forms. Unpublished doctor's dissertation, The Ohio State Univ., 1954.
18. WEINSTEIN, M. Stimulus complexity and the recognition of visual patterns. Unpublished doctor's dissertation, The Ohio State Univ., 1955.
19. WERTHEIMER, W. *Productive thinking.* New York: Harper, 1945.

(Received for early publication
June 17, 1955)

Journal of Experimental Psychology
Vol. 51, No. 1, 1956

RELATIONSHIP BETWEEN RECOGNITION ACCURACY AND ORDER OF REPORTING STIMULUS DIMENSIONS[1]

DOUGLAS H. LAWRENCE AND DAVID L. LABERGE[2]

Stanford University

Since the studies of Külpe, Grunbaum, and others (3), it has been known that when heterogeneous stimulus objects are presented tachistoscopically to an *S* his accuracy of report for stimulus aspects he has been instructed to observe is greater than for other incidental aspects (cf. 1). For instance if the stimulus objects differ in color, form, and spatial location, *S* can readily report the colors if he has been instructed to do so prior to the presentation of the stimulus, but he may have little if any knowledge of form and spatial position. Külpe accounted for this selective influence of instructions in terms of two perceptual processes: (*a*) positive abstraction, which accentuated the crucial aspect, and (*b*) negative abstraction, which actively inhibited the perception of the incidental aspects. This way of formulating the problem seems to imply that these processes operate at the moment of stimulus reception and that they achieve their result by filtering out a large part of the stimulus input so that the incidental aspects never are available to *S*.

Assuming this is the correct interpretation of Külpe's hypothesis, it raises several questions about the selective nature of perception. First of all it is not clear from this formulation whether the over-all accuracy of report for all aspects of the stimulus should increase, decrease, or remain constant under these special instructions as contrasted to cases in which no such instructions are given. It can be argued that because the stimulus contains more incidental than crucial aspects, the inhibition of the former would reduce the over-all accuracy of report, i.e., positive abstraction of one aspect could not compensate for the negative abstraction of several. In information terms the increase in transmitted information about the crucial aspect would not compensate for the loss in transmitted information about the incidental aspects. Secondly, it is not clear that the selective processes postulated by Külpe are necessarily operating at the time of stimulus reception. A previous study (4) on the accuracy of recognition suggests that task sets have similar influences on recognition when given after as when given prior to the stimulus presentation. This would indicate that their influence is on memory and response factors rather than on stimulus reception.

This experiment attempts a partial answer to these questions concerning Külpe's results. One group of *S*s is instructed to pay equal attention to all aspects of a complex stimulus, whereas a second group is instructed to give primary attention to the crucial aspect only. Afterward, however, both groups report on all aspects. This essentially duplicates Külpe's situation and at the same time makes

[1] This experiment is one in a series of studies of problem solving being done under Project NR 150–149 and supported by Contract N6onr 25125 between Stanford University and the Office of Naval Research. Work on the contract is under the general direction of Dr. Donald W. Taylor. Permission is granted for reproduction, translation, publication, use, and disposal of this article in whole or in part by or for the United States Government.

[2] Now at Indiana University.

possible a comparison of the two sets of instructions for the over-all accuracy of report. A third group has instructions to pay attention to and to report only the crucial dimension to see if the positive abstraction effect can be further enhanced. A fourth group is instructed to pay equal attention to all aspects, but immediately after the presentation of the stimulus, *E* specifies the order in which the aspects are to be reported. If this group shows the same difference in accuracy of recognition between the first aspect reported and the others as between the crucial and incidental aspects in the first group, this would suggest that the positive and negative abstraction effects are largely due to memory and response factors.

Procedure

The stimuli for this experiment were constructed from the cards in the Wisconsin Card Sorting Task (2). These cards differ in color (yellow, green, red, and blue), form (circles, stars, triangles, and crosses), and number of figures per card (one, two, three, and four). Two of these cards were placed side by side to form a stimulus. With the restriction that the two cards of a stimulus should differ in color, form, and number, there are 216 possible pairings. These 216 stimuli were divided into four sets of 54, and each of these into three subsets of 18. Each subset of 18 contained three replications of the six pairings of color, of form, and of number. Each value of a stimulus dimension, such as the color yellow, occurred on the right card and the left card of a pair an equal number of times.

The tachistoscope consisted of an opaque projector with a camera shutter mounted in front of the lens. Each stimulus was exposed for 1/10 sec. on a screen 6 ft. in front of the *Ss*. The room was dimly lighted so that they could record their responses on prepared forms during the 15-sec. interval between exposures.

The four different types of instructions used in this experiment were as follows:

Equal instructions.—The *Ss* were instructed to pay equal attention to all three dimensions, color, form, and number, and to record all three even though they might have to guess. They were told to imagine that they were being paid $34 for each dimension correctly reported. They were permitted to record their answers in any order.

Emphasis instructions.—The *Ss* were instructed to pay primary attention to one dimension (dominant) and to ignore the other two (nondominant), but nonetheless to record all three in any order even if they had to guess. They were told to imagine that they were being paid $100 for each correct report about the dominant but only $1 for each nondominant dimension.

One-only instructions.—These were exactly the same as the emphasis instructions with the exception that *Ss* were told to imagine that nothing would be paid for the nondominant dimensions. Consequently they recorded only the dominant one.

Ordered instructions.—These were exactly the same as the equal instructions. After the presentation of each stimulus, however, *E* specified the order in which the stimulus dimensions were to be recorded, e.g., color first, number second, and then form. The six possible orders were used equally often and occurred in random sequence so that *Ss* could not anticipate them.

Because of a scarcity of *Ss* at the time the study was made, a somewhat unorthodox design was used. Group I, consisting of 40 Stanford undergraduates, was divided into four subgroups of 10. Each subgroup had a set of 54 stimuli under Equal and a second set of 54 under Emphasis instructions. Half the subgroups had Equal first followed by Emphasis, and half the converse order. All four sets of stimuli occurred under each type of instructions. Under Emphasis instructions color was the dominant cue for a subset of 18 stimuli, form was dominant for a second subset, and number for a third. The order in which these dimensions were dominant varied between subgroups.

Group II, consisting of 20 Stanford undergraduates run at a later time, was divided into four subgroups of five. These *Ss* had a set of 54 stimuli under Equal and a second set under Ordered instructions with order of instructions balanced between subgroups. All four sets of stimuli occurred under each type of instructions. By having this group replicate the Equal instructions it was possible to determine its comparability with Group I.

Group III, consisting of 20 City College of San Francisco undergraduates, was divided into four subgroups of five. Each subgroup had a set of 54 stimuli under Emphasis and second set under One-only instructions with order of instructions balanced between subgroups. Under One-only instructions the ordering of the dominant dimension was manipulated in the same way as for Emphasis instructions. By having this group replicate the Emphasis instructions it was

TABLE 1

AVERAGE ACCURACY OF REPORT PER SUBSET OF 18 STIMULI
(Maximum possible score = 36: Chance score = 18)

Group	Instructions		Stimulus Dimension			Mean	% Maximum
			N	C	F		
I	Emphasis	Dominant	31.50	29.34	32.76	31.20	86.7
		Nondominant	26.04	27.12	26.04	26.40	73.3
		Weighted Average	27.84	27.84	28.26	28.02	77.8
	Equal		29.88	28.02	29.16	29.04	80.7
II	Equal		29.88	26.88	28.98	28.56	79.3
	Ordered	1	31.02	28.50	32.10	30.54	84.8
		2	27.72	27.36	27.30	27.48	76.3
		3	27.60	25.56	26.34	26.52	73.7
		Average	28.80	27.12	28.56	28.20	78.3
III	One-only		32.70	27.24	34.02	31.32	87.0
	Emphasis	Dominant	31.32	27.90	33.06	30.78	85.5
		Nondominant	24.68	24.96	26.16	25.26	70.2
		Weighted Average	26.88	25.92	28.44	27.12	75.3

possible to determine its comparability with Group I.

All *Ss* were run in subgroups of five. They were given preliminary practice during which they were shown samples of all three stimulus dimensions and of all four values on a dimension. They were told that the two cards forming a stimulus never had exactly the same values on a given dimension but that otherwise all six possible pairings would occur. They were then presented with five stimuli for the regulation exposure time, they recorded their answers, and these were corrected after each exposure. No testing for color blindness was done.

RESULTS

Each *S* was assigned three scores on each subset of 18 stimuli, one for color, one for form, and one for number. As each stimulus contained two values on each dimension, e.g., blue and yellow for color, *S* could report 0, 1, or 2 correctly. Thus his maximum possible score per dimension on 18 stimuli was 36, his chance score 18. The average accuracy for each group per dimension under the various types of instruction is shown in Table 1. The statistical significance of these results was determined by an analysis of variance for each group separately. These are shown in Tables 2, 3, and 4. In these analyses all the interactions between

TABLE 2

ANALYSIS OF VARIANCE FOR GROUP I ON ACCURACY OF RECOGNITION

Component	*df*	Mean Square	*F*
1. Stimulus dimensions	2	60.58	8.51**
2. Instructions	1	188.09	26.42**
3. Subsets within instructions			
a. Dominant vs. nondominant	1	1836.81	257.98**
b. Between nondominant	1	0.42	—
c. Within equal	2	65.18	9.15**
4. Interactions between conditions			
a. 1 × 2	2	54.65	7.68**
b. 1 × 3a	2	143.94	20.22**
c. 1 × 3b	2	50.75	7.13**
d. 1 × 3c	4	1.07	—
5. Subgroups	3	14.57	
6. Individuals	36	86.82	
7. Interactions with subgroups	51	12.45	
8. Interactions with individuals†	612	7.12	

** .01 level of significance.
† Error term used throughout.

subgroups and each of the other components were computed as were the interactions between individuals and each component. In each case tests for homogeneity of variance did not reach the .01 level of significance. Consequently only the weighted averages of these two types of interactions are presented in the tables.

The important findings for Group I are as follows. (*a*) The over-all level of accuracy of report for the three dimensions combined is slightly greater under Equal than under Emphasis instructions (80.7% vs. 77.8%). This difference, though small, is significant at the .01 level. (*b*) Under Emphasis instructions the accuracy of report is considerably greater on dominant dimensions than on nondominant ones (86.7% vs. 73.3%). This difference is significant well beyond the .01 level. These two findings indicate that the drop in over-all accuracy of report from Equal to Emphasis instructions is very slight, and that the results of Külpe's experiment have been duplicated under the present conditions.

TABLE 3

ANALYSIS OF VARIANCE FOR GROUP II ON ACCURACY OF RECOGNITION

Component	*df*	Mean Square	*F*
1. Stimulus dimensions	2	175.69	25.28**
2. Instructions	1	15.09	2.17
3. Subsets within instructions			
a. 1st vs. 2nd & 3rd	1	511.22	73.56**
b. 2nd vs. 3rd	1	27.08	3.90*
c. Within equal	2	11.71	1.68
4. Interaction between conditions			
a. 1 × 2	2	13.00	1.87
b. 1 × 3a	2	34.98	5.03**
c. 1 × 3b	2	7.22	1.04
d. 1 × 3c	4	4.86	—
5. Subgroups	3	75.50	
6. Individuals	16	68.00	
7. Interaction with subgroups	51	10.55	
8. Interaction with individual†	272	6.95	

* .05 level of significance.
** .01 level of significance.
† Error term used throughout.

TABLE 4

ANALYSIS OF VARIANCE FOR GROUP III ON ACCURACY OF RECOGNITION

Component	*df*	Mean Square	*F*
1. Stimulus dimensions	2	256.13	32.26**
2. Between conditions			
a. One-only vs. dominant	1	9.08	1.14
b. One-only and dominant vs. nondominant	1	2001.04	252.02**
3. Between replications of nondominant†	3	33.72	4.25**
4. Interaction between conditions			
a. 1 × 2*a*	2	11.20	1.41
b. 1 × 2*b*	2	151.22	19.05**
5. Subgroups	3	56.12	
6. Individuals	16	71.03	
7. Interaction with subgroups	33	9.44	
8. Interaction with individual‡	176	7.94	

** .01 level of significance.
† This component is confounded with the interaction between stimulus dimensions and replications of the nondominant stimuli.
‡ Error term used throughout.

Other subsidiary findings for Group I are as follows. On the average, accuracy of report for color is poorer than for the other two dimensions. This difference is larger under Equal than under Emphasis instructions with the result that there is a significant interaction between instructions and stimulus dimensions. Similarly, under Emphasis instructions, color fails to show as large a difference in accuracy between the dominant and nondominant dimensions as do form and number. This also produces a significant interaction effect. The significant differences found between subsets of stimuli under Equal instructions is due to a small but consistent practice effect.

The comparability of Group II and Group I *S*s was determined by their respective performances under Equal instructions. Each *S* was assigned

three scores, one on each of the three stimulus dimensions. None of the three *t* tests was significant at the .10 level.

The important findings for these Group II *S*s are as follows. (*a*) The difference in over-all accuracy of report under Equal and Ordered instructions is small and nonsignificant (79.3% vs. 78.3%). (*b*) Under Ordered instructions each *S* was assigned three scores, one for the first stimulus dimension reported, one for the second, and one for the third. The difference in accuracy of report on the first and on the average of the second and the third is large and significant beyond the .01 level (84.8% vs. 75.0%). It is of the same magnitude as the difference between dominant and nondominant dimensions that was found for Group I *S*s under Emphasis instructions. There is a small difference in accuracy between the second and third dimensions reported which is significant at the .05 level. These findings indicate that the Külpe effect can be duplicated just by the order in which the responses are made even though the *S*s are set to give equal attention to all three stimulus dimensions.

Subsidiary findings for this group are as follows. There is again a significant difference between stimulus dimensions largely due to the greater difficulty in reporting color. This dimension fails to show as great a difference in accuracy under Ordered instructions between the first and second reported dimensions as do form and number. This produces a significant interaction which is directly comparable to what was found for Group I under Emphasis instructions.

The comparability of Group III and Group I *S*s was determined by their respective performances under Emphasis instructions. Each *S* was assigned two scores, one for the combined dominant dimensions and one for the combined nondominant dimensions. Neither *t* test approached significance at the .10 level.

It was impossible to compare the over-all accuracy of this group under One-only and Emphasis instructions as the *S*s recorded only one dimension per stimulus under the former. The important findings for this group are as follows. (*a*) The average accuracy under One-only is only slightly and nonsignificantly different from that on the dominant dimension under Emphasis instructions (87.0% vs. 85.5%). (*b*) When the average accuracies for One-only instructions and for the dominant dimension under Emphasis instructions are combined and contrasted with that on the nondominant dimensions under Emphasis instructions, the difference is large and significant well beyond the .01 level (86.3% vs. 70.2%). These findings indicate that having to report the two nondominant dimensions under Emphasis instructions seemingly has not interfered with accuracy on the dominant dimension. It suggests that *S*s have more than a chance amount of information about the incidental or nondominant aspects of a stimulus even when they have been instructed to completely ignore them. In this group there is additional evidence that color is more poorly reported than the other stimulus dimensions and that it results in an interaction with the dominant versus nondominant dimensions under Emphasis instructions.

Discussion

The results of this experiment indicate that the Külpe effect can be consistently shown under the present conditions. On the other hand, they throw considerable doubt on Külpe's interpretation of these as due to positive and negative abstrac-

tion processes of a perceptual nature that occur during stimulus reception.

The fact that under three different types of instruction, i.e., Equal, Emphasis, and Ordered, the over-all average accuracy differs but little suggests that the transmitted information under tachistoscopic presentation tends to be constant irrespective of instructions. The manner in which this effective information is distributed between the dimensions, however, can be influenced by instructions. Under Emphasis instructions the increased accuracy on the dominant dimension is just offset by the losses on the nondominant dimensions. It would appear that accuracy was maximal on the dominant dimension for the stimuli and exposure time used in that it could not be increased by the One-only instructions. Thus the fact that *S*s had to report the nondominant dimensions under Emphasis instructions, even though they had been told to pay attention to the dominant dimension only, was not an interfering factor.

The results for the Ordered instructions strongly suggest that the Külpe effect is largely due to memory and response factors rather than to perceptual ones at the time of stimulus reception. The *S*s under these instructions, even though told prior to stimulus presentation to pay equal attention to all three dimensions, actually duplicated the results for the Emphasis instructions. They could not have been anticipating the order in which the dimensions were to be recorded because this varied randomly from trial to trial and was not specified until after the stimulus had been exposed. Consequently it must have been the act of writing down the first dimension plus the increased temporal delay that interfered with the accuracy of report for the second and third dimensions. Thus the order of accuracy under these instructions must be accounted for in terms of memory and response factors.

If it is assumed that *S*s under Emphasis instructions had a strong tendency to report the dominant dimension first, these same memory and response factors would account for their results. Unfortunately no record was kept for these instructions of the actual order in which the dimensions were written down by each *S*. The *E*'s observations and the comments of *S*s after the experiment, however, indicate that the dominant dimension was recorded first. Similarly if it is assumed that under Equal instructions the order in which dimensions were recorded varied between *S*s and for the same *S* at different times, the order effect would cancel out when averaged over all trials and *S*s. This would give just the results found for these instructions. Again the only available evidence is *E*'s observations and the comments of *S*s, both of which support this assumption. Thus all the results of this experiment can be accounted for by assuming: (*a*) that under tachistoscopic presentation a constant amount of information is transmitted irrespective of instructions, and (*b*) that how this information is distributed between dimensions is largely a function of the order in which they are reported. This would mean that the selective effect of instructions operates primarily through memory and response factors.

Summary

Stimulus objects differing in color, form, and number of objects were presented tachistoscopically to groups of *S*s. Four types of instruction were given: (*a*) Equal instructions in which *S*s were asked to pay equal attention to and to record all three dimensions, (*b*) Emphasis instructions in which *S*s were asked to pay primary attention to one dimension only but to record all three, (*c*) One-only instructions in which *S*s were asked to pay primary attention to and to record only one dimension, and (*d*) Ordered instructions in which *S*s were asked to pay equal attention to and to record all three dimensions but the order in which they were recorded was specified immediately after the stimulus exposure. The main findings and conclusions were as follows:

1. The average accuracy of report on all three dimensions combined differed but slightly between the various types of instruction.

2. Under Emphasis instructions the accuracy of report on the dimension *S*s had been instructed to pay attention to was considerably greater than on the other two dimensions.

3. This accuracy on the dominant dimension was not increased under One-only instructions even though *Ss* did not report the other two dimensions.

4. Under Ordered instructions the difference in accuracy between the first recorded dimension and the average of the other two was as large as the difference reported for Emphasis instructions.

5. It was concluded that the selective effect of instructions on the accuracy of report for tachistoscopically presented stimuli can be accounted for in large part by the following assumptions: (*a*) a constant amount of information is transmitted irrespective of instructions; (*b*) the distribution of this between dimensions depends upon the order in which they are reported; and (*c*) the effect of instructions is to determine which order of report occurs.

References

1. Glanville, A. D., & Dallenbach, K. M. The range of attention. *Amer. J. Psychol.*, 1929, **41**, 207–236.
2. Grant, D. A., & Curran, J. F. Relative difficulty of number, form, and color concepts of a Weigl-type problem using unsystematic number cards. *J. exp. Psychol.*, 1952, **43**, 408–413.
3. Humphrey, G. *Thinking: an introduction to its experimental psychology*. New York: Wiley, 1951.
4. Lawrence, D. H., & Coles, G. R. Accuracy of recognition with alternatives before and after the stimulus. *J. exp. Psychol.*, 1954, **47**, 208–214.

(Received January 25, 1955)

Journal of Experimental Psychology
Vol. 51, No. 1, 1956

COLOR CODING IN A VISUAL SEARCH TASK [1]

BERT F. GREEN AND LOIS K. ANDERSON

Lincoln Laboratory, Massachusetts Institute of Technology

Complex visual displays are used to present information in many information-processing systems. The operator's task often involves locating a particular symbol or group of symbols on the display. Since the time required to perform the search task is a critical factor in such systems, methods are frequently devised for shortening the search times. One method is to improve the legibility of the symbols. Another method is to make additional distinctions among the symbols by using, for example, symbols of different colors or sizes. Such additional distinctions, or codings, are redundant in the sense that the operator does not need them in order to find particular targets. However, if the operator knows the color, or size, of the target, he may be able to find it much quicker. The effectiveness of a redundant code might be measured by the reduction in search times when the code is used.

Green, McGill and Jenkins (**4**), using three-digit numbers as symbols, found that the average time required to find a particular target number was nearly proportional to the number, or density, of symbols. When they used a color code in which half of the numbers were blue and half were yellow, *O*s who were told the color of the target number found the target in about half the time required by *O*s who were not told the color. In this instance, color coding seemed to be completely effective in reducing the number of alternatives through which *O* searched. The purpose of the present experiments was to investigate the effectiveness of color coding as a function of the relative number of symbols of each color, and as a function of the number of different colors used. We were interested in the latter variable because Eriksen (**2**, **3**) had found that search times depend, in a rather complicated way, on the heterogeneity of the visual field.

[1] This research was supported jointly by the Department of the Army, the Department of Navy, and the Department of the Air Force under Air Force Contract No. AF 19 (122)-458.

Experiment I

The purpose of the first experiment was to measure the effectiveness of a two-color code as a function of the relative number of symbols of each color. The results of Green, McGill, and Jenkins (**4**) suggested that when *O* was told the target color, the search times would be proportional to the number of symbols of the same color as the target symbol, but when *O* was not told the target color, the search times would be proportional to the total number of symbols.

Procedure.—Our displays contained two-digit numbers in the range 10-69, arranged in random order in a matrix of 10 rows and 6 columns. They were prepared as Kodachrome transparencies and projected onto a screen 10 ft. from *O*. The projected image of the matrix was 16¾ in. high and 12 in. wide, with numerals 1¼ in. by ¾ in. The numbers were either green or red, corresponding approximately with Munsell 5.0 GY/6/6 and 2.5 YR/6/10, respectively, and were presented on a black background.

Three experimental conditions were used. In the Set condition we used displays containing 60 numbers, of which 0, 10, 20, 30, 40, 50, or 60 were red, and the others green. No number appeared more than once on the display so that color was not essential to the search task. At the beginning of each trial in the Set condition, *E* announced the target number and its color. When *O* had repeated the number, *E* opened the

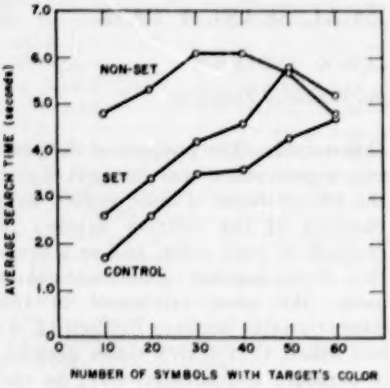

FIG. 1. Average search time for colored numbers in Exp. I. Each point represents the geometric mean of 80 measurements. All Non-set and Set displays had 60 symbols; Control displays had from 10 to 60 symbols, all of the same color.

projector shutter and started a stop watch. As soon as *O* found the number, he said "Now" which signaled *E* to stop the watch. The *O* then used a special flashlight pointer to indicate the target on the screen, so that *E* could check for errors. The same procedure and the same displays were used in the Non-set condition with the exception that *O* was not told the color of the target number. In the Control condition the displays had 10, 20, 30, 40, 50, or 60 numbers of one color, either red or green, while the remaining positions in the matrix were blank. The *E* announced the target number at the beginning of each Control trial, which then proceeded like the Set and Non-set trials.

The *O*s were 20 U. S. Air Force enlisted men with normal visual acuity, as measured by the Orthorater, and normal color vision as measured by the American Optical Company's pseudoisochromatic plates. Each *O* was given six practice trials on each condition, and two experimental trials on all combinations of three conditions, two target colors, and six densities of symbols of the target's color. In the experimental design, we also balanced the order of presentation of conditions, the order of trials within conditions, the target color, and the particular number and matrix position used as the target.

Results.—The average search times for each condition in Exp. I are shown in Fig. 1. If the only effect of color coding were to reduce the number of alternatives through which *O* must search, the Non-set curve would be flat, because in this condition there were always 60 symbols on the display and *O* did not know the target's color.

TABLE 1

ANALYSIS OF VARIANCE FOR EXP. I

Source	Set and Control Data			Non-Set Data		
	df	Mean Square	*F*	*df*	Mean Square	*F*
Observers	19	.291	4.69***	19	.383	4.45***
Hues	1	.025	—	1	.508	5.91*
Conditions	1	2.719	43.85***			
Densities	5	3.342	53.90***	5	.178	2.07
O × H	19	.045	—	19	.092	1.07
O × C	19	.101	1.63*			
O × D	95	.060	—	95	.105	1.22
H × C	1	.093	1.50			
H × D	5	.082	1.32	5	.050	—
C × D	5	.087	1.40			
O × H × C	19	.031	—			
O × H × D	95	.057	—	95	.096	1.12
O × C × D	95	.062	1.00			
H × C × D	5	.039	—			
O × H × C × D	95	.057	—			
Replications	480	.062		240	.086	
Total	959			479		

* $P = .05$.
*** $P = .001$.

Actually, the Non-set curve has a maximum near 30–40 symbols and decreases slightly on either side of this maximum. An analysis of variance of the Non-set data, shown in Table 1, indicates that the main effect of density is not significant. (Since the means of the search time distributions were roughly proportional to their *SD*'s, a logarithmic transformation was used. This transformation did not completely overcome the correlation of means and *SD*'s so the significance levels in the analyses should be interpreted conservatively.) We made a regression analysis of density by partitioning the main effect of density into linear, quadratic, and residual components. The quadratic mean square was much larger than the other two, and was significant at the .01 level, which indicates that the Non-set curve can be closely approximated by a quadratic regression on density. Thus, the regression analysis indicates that the Non-set curve is significantly bowed while the standard analysis of variance indicates that the curve is flat. The reason for the contradiction seems to be that the regression analysis takes account of the logical order of the density values, while the standard analysis of variance does not. The regression analysis suggests that displays on which one color predominates are slightly easier to search than displays on which the two colors are more evenly represented.

If the mean search times were proportional to the number of alternative symbols, then the Control curve would be a straight line through the origin. It appears that an extrapolation of the Control curve would intercept the *y* axis (zero density) at about 1 sec. This is essentially the origin for the data, since the reaction times involved in making a measurement add to about 1 sec. This is indicated by the fact that we recorded several measurements of 1.0 sec., but none below this value. With this correction, the Control curve is closely proportional to density, as we expected. All three curves should meet at the target density of 60, since all conditions used equivalent displays at this density, i.e., 60 numbers of a single color. The observed differences must be attributed to chance.

If *O* could completely disregard symbols of a different color from the target, then the Set curve would be identical with the Control curve. That is, wrong-colored symbols would have the same effect as blanks in the matrix. Actually the average search times in the Set condition are slightly longer than the corresponding times in the Control condition. The statistical significance of this difference is established by the significant main effect of conditions in the analysis of variance of the Set and Control data, as shown in Table 1. Since the Set curve rises regularly, we may conclude that the average search times in the Set condition depend mainly on the number of symbols with the target's color, but are also affected slightly by the presence of wrong-colored symbols.

Experiment II

The difference between the Set and Control conditions in Exp. I is consistent with Eriksen's results (**2**, **3**) showing that display heterogeneity is an important factor in search times. These results led us to measure search times for displays with varying numbers of colors.

Procedure.—The displays were similar to those used in Exp. I but they included red, yellow, green, and blue numbers. The yellow and blue colors corresponded roughly with Munsell 2.5 Y/8/10 and 10G/7/4, respectively. In the Control condition we used displays with 15, 30, 45, or 60 numbers of a single color, either red, yellow, green, or blue. The procedure was otherwise

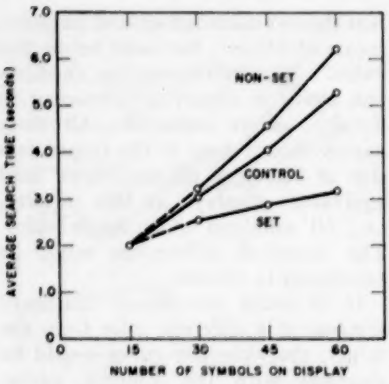

FIG. 2. Average search times for colored numbers in Exp. II. Each point represents the geometric mean of 240 measurements. Non-set and Set displays had 15 symbols each of two, three, or four colors; all symbols on Control displays had the same color.

identical with that of Exp. I. In the Set condition we used displays containing 30, 45, or 60 numbers that were 15 each of two, three, or four colors, respectively. All possible combinations of colors were represented. In the Set condition, *O* was told the color of the target; there were always 15 numbers of the target's color, but there might also be 15 numbers of one wrong color and 30 blanks, or 15 numbers of each of two wrong colors and 15 blanks, or 15 numbers of each of three wrong colors. The displays used in the Non-set condition were the same as those in the Set condition, but *O* was not told the color of the target. (We omitted the density of 15 from the Set and Non-set conditions since it was logically equivalent to the density of 15 in the Control condition.)

The *O*s were 20 U. S. Air Force enlisted men with normal visual acuity and normal color vision. Each *O* was given four practice trials on each condition, and three experimental trials on each of the four colors for each of the 10 density-condition combinations. The experimental design was otherwise similar to that used in Exp. I.

Results.—The average mean search time for each density-condition combination is shown in Fig. 2. The Set curve increases slightly as the number of symbols on the display increases, although there were always 15 symbols of the target's color. This is consistent with the difference between

TABLE 2

ANALYSIS OF VARIANCE FOR EXP. II

Source	Non-Set and Control Data			Set Data		
	df	Mean Square	*F*	*df*	Mean Square	*F*
Observers	19	.503	5.65***	19	.227	5.68***
Hues	3	.027	—	3	.145	3.62*
Conditions	1	.860	9.66**			
Densities	2	8.198	92.11***	3	2.134	53.35***
O × H	57	.089	1.00	57	.039	—
O × C	19	.180	2.02**			
O × D	38	.078	—	57	.053	1.32
H × C	3	.081	—			
H × D	6	.074	—	9	.082	2.05*
C × D	2	.066	—			
O × H × C	57	.077	—			
O × H × D	114	.075	—	171	.046	1.15
O × C × D	38	.075	—			
H × C × D	6	.096	1.08			
O × H × C × D	114	.077	—			
Replications	960	.089		640	.040	
Total	1439			959		

* $P = .05$.
** $P = .01$.
*** $P = .001$.

Set and Control curves in Exp. I, and indicates that the presence of wrong-colored symbols has a small effect on search times in the Set condition. In Exp. II, the number of different colors on the display increased as the number of wrong-colored symbols increased, so the two effects are confounded in the Set condition. The significance of the main effect of density in the analysis of variance of the Set data, shown in Table 2, shows that the increase in the Set curve is statistically significant. (The data for density 15 were analyzed with the Set data rather than with the Control data.)

The Control curve is closely proportional to symbol density, as it was in Exp. I, provided the 1-sec. correction is made. The Non-set curve is slightly above the Control curve. This difference is supported by the significance of the main effect of conditions in the analysis of variance of the Non-set and Control data, shown in Table 2. The difference between the Non-set and Control curves indicates that search times are longer for displays with equal numbers of symbols of several colors than for displays with comparable numbers of symbols all of the same color. We found the same effect in Exp. I where it was evidenced by the bowed shape of the Non-set curve.

Discussion

When *O* knows the target's color, his average search time depends primarily on the number of symbols of the same color as the target. Symbols of other colors have only a minor effect on the search time. Although the information provided by the color code is redundant, or nonessential, it is clearly useful. On the other hand, the use of a redundant color code involves some loss if *O* does not know the code category of the target, since search times are longer for multi-colored than for single-colored displays. The latter result is somewhat surprising since one might expect that search time would be inversely related to the distinctiveness of the symbols, and that using several colors would add to the distinctiveness of the symbols. In psychophysical terms, we might consider "signal" to be those aspects of the display that aid in locating the target, "noise" to be those aspects of the display that detract from target location, and "signal-to-noise ratio" to be analogous (inversely) to search time. Then our results would show that the redundant color code adds to the signal under the Set condition, but adds to the noise under the Non-set condition.

Since almost any visual code can be tested in a visual search task, we suggest that codes can be evaluated in terms of search times. We would expect search times to be sensitive to such factors as the discriminability of the coded symbols, the difficulty in learning the code, etc. The evaluation should consider both the reduction in search times in the Set condition and the increase in search times in the Non-set condition, for the relative importance of these factors depends on how often, in a particular operational situation, *O* will know the target's code category and how often he will not.

Color could be considered irrelevant in our Non-set condition. Recently, the term *irrelevant* has been used differently by some authors for whom irrelevant information is additional stimulus information not correlated with the correct response. Gregg (5) increased the amount of irrelevant information in a binary stimulus situation, and found an increase in the time required to make a binary discriminative response. Archer (1) used stimuli with various amounts of relevant and irrelevant information, and found no difference in response times in an identification task for different amounts of irrelevant information. In our experiment, color was not irrelevant in their sense of the term, since color did not increase the stimulus information, but was always redundant. We used

verbal instructions to render color relevant to *O* in the Set condition and irrelevant to *O* in the Non-set condition. Obviously, instructions as well as the structure of the stimulus are important in defining relevance.

Our experiments are similar to many others that have dealt with set as a selective factor in perception. Our results are consistent with the general finding that simple, or "single," sets, as Postman and Bruner (7) would call them, have clear, simple selective effects (for example, the classic experiment of Külpe and Bryan (6)). It would be interesting to apply other findings from this area to our specific problem. For example, Postman and Bruner studied multiple sets, in which *S* was not told the precise category of a target, but was told that the target might be in one of two or more specified mutually exclusive categories. They found that *S*s with multiple sets did not perform as well as would have been predicted from single set data. This implies that in our four-color display, if *O*s knew only that the target was, say, either red or green, but not blue or yellow, the search times would be more than twice as long as the times for single color sets.

Summary

Two experiments were reported in which search times for colored symbols (two-digit numbers) on a visual display were measured as a function of the relative number of symbols of each color, and the number of different colors used. When *O*s know the color of the target, the search time is approximately proportional to the number of symbols of the target's color. There is also a slight increment in search time due to the presence of the wrong-colored targets. When *O*s do not know the target's color, search time depends primarily on the total number of symbols on the display. However, search times are slightly longer for multicolored displays than for comparable single-colored displays.

References

1. Archer, E. J. Identification of visual patterns as a function of informational load. *J. exp. Psychol.*, 1954, **48**, 313–317.
2. Eriksen, C. W. Location of objects in a visual display as a function of the number of dimensions on which the objects differ. *J. exp. Psychol.*, 1952, **44**, 56–60.
3. Eriksen, C. W. Object location in a complex perceptual field. *J. exp. Psychol.*, 1953, **45**, 126–132.
4. Green, B. F., McGill, W. J., & Jenkins, H. M. The time required to search for numbers on large visual displays. Massachusetts Institute of Technology: Lincoln Laboratory, Technical Report No. 36, August 18, 1953.
5. Gregg, L. W. The effect of stimulus complexity on discriminative responses. *J. exp. Psychol.*, 1954, **48**, 289–297.
6. Külpe, O., & Bryan, W. L. Versuche über Abstraktion. *Ber. I Kongr. exp. Psychol.*, 1904, 58–68.
7. Postman, L., & Bruner, J. S. Multiplicity of set as a determinant of perceptual behavior. *J. exp. Psychol.*, 1949, **39**, 369–377.

(Received January 28, 1955)

Journal of Experimental Psychology
Vol. 51, No. 1, 1956

TIME CONTINUOUSLY ON TARGET AS A FUNCTION OF DISTRIBUTION OF PRACTICE [1]

LYLE E. BOURNE, JR. AND E. JAMES ARCHER

University of Wisconsin

The most commonly used method of measuring rotary pursuit performance has been to connect a clock into the stylus-target circuit and to record the total time of circuit completion for each trial. This record, however, gives no information as to the patterning of the times continuously on target, i.e., "hits," during the trial. The frequency distribution of durations of time continuously on target offers a different and more analytic performance measure for the investigation of motor skills. Several phenomena might be examined further using this measure. Improvement with practice, facilitation due to distributed practice, and reminiscence may all take the form of (*a*) a simple increase in number of hits, (*b*) a shift from short to long hits, or (*c*) a combination of both effects. There have been two attempts to analyze motor performance in terms of duration of hits. Ammons (**5**), using a kymographic recording technique, investigated the change in duration of stylus-target circuit completion times in a pursuit rotor task. Because of the inordinate amount of time required to read individual records (12 to 15 hr. per record), only a small number of *S*s was used. The results are not as informative as might be expected since only mean number of hits and mean duration of hits per trial were reported. Archer, Kent, and Mote (**7**) using a Tracking-time Analyzer, a device which electronically measures frequency of times continuously on target, have determined the effect of long-term practice on performance in the Pedestal Sight Manipulation Test (PSMT). These *E*s found that as practice progressed there was an increase in the duration of hits and the total number of hits first decreased and then increased. The time-continuously-on-target analysis, applied to a simpler motor task such as rotary pursuit, could add to the information now available about the effects of distributed practice.

Kimble (**10**) posits a two-factor theory of inhibition which holds that the total inhibitory potential generated during practice is accounted for by reactive inhibition, I_R, and conditioned inhibition, ${}_sI_R$. Both components are of course detrimental to performance. Reactive inhibition is considered a negative drive which develops from response evocation and dissipates during rest, whereas ${}_sI_R$ is considered a habit which is conditioned to stimuli associated with resting and does not dissipate spontaneously. According to Kimble, the amount of I_R generated in a number of responses can be determined by the amount of reminiscence shown after a suitable rest interval. The difference between prerest distributed- and massed-practice groups in initial postrest performance is taken as an indication of ${}_sI_R$. Most commonly, time-on-target scores have been used in obtaining the measures of the two inhibitory potentials. Using times continuously on target as a performance measure, additional information

[1] Supported in part by a grant from the Research Committee from funds provided by the Wisconsin Alumni Research Foundation.

related to the existence of a permanent work decrement, or conditioned inhibition in motor performance could be obtained.

Kimble (11) and Wasserman (15), using an inverted alphabet printing task, and Ammons (4), Barch (8) and Kimble and Shatel (12), using a rotary pursuit task, found that massed-practice groups, after an interpolated rest, showed a significant increase in performance (reminiscence) but remained significantly poorer in postrest performance than a distributed-practice group. They explained the reminiscence in terms of the dissipation of I_R or temporary work decrement and the significant postrest difference between groups in terms of ${}_sI_R$ or permanent work decrement. Adams and Reynolds (1), Archer (6), Schucker, Stevens, and Ellis (13), and Starkweather and Duncan (14) have found no significant effect of permanent work decrement in motor performance. Archer believes that the difference between massed and distributed groups found by Kimble is a chance phenomenon. He criticizes Kimble for comparing the performance of the massed group after 10-min. rest with the performance of the distributed group after only 30-sec. rest and for using individual t tests without first establishing a true difference between groups with an over-all analysis of variance. Schucker, Stevens, and Ellis suggest that the differences Kimble obtained were due to uncontrolled factors in his experiment such as rest-interval activity, knowledge of results, and number of responses evoked at the points of comparison. The data of the present experiment, recorded in time continuously on target, may help to clarify the controversy which has resulted from previous experiments.

Procedure

Subjects.—One hundred women students in introductory psychology courses at the University of Wisconsin served as *Ss* in this experiment. These *Ss*, naive to the apparatus and task, were assigned randomly to one of five groups and served individually for one session.

Apparatus and task.—The task was provided by a 54-rpm pursuit rotor. The turntable had a radius of 6 in. The target was a silver disc of .75 in. diameter, set flush with the surface of the turntable; the center of the disc was 5 in. from the center of the turntable. The *S* tracked the target with a stylus which was hinged to prevent her from pressing down on the target. When the stylus point contacted the target, a circuit was completed to the Tracking-time Analyzer. The Tracking-time Analyzer combines a bank of 12 resettable Mercury counters and a .01-sec. scale Standard Electric clock. The clock was used to obtain the conventional record of total time on target. The counters recorded the number of hits of various durations continuously on target. For each counter except the first there was a separate thyratron-condenser-resistor circuit having its own critical delay period. The first counter, activated by a direct impulse from the stylus-target contact, recorded all hits regardless of duration. The second counter recorded all hits longer than .05 sec. in duration; the third recorded hits longer than .10 sec. The other intervals were .20, .30, .40, .60, .80, 1.00, 1.20, 1.60, and 2.00 sec. With the Tracking-time Analyzer it is possible to analyze total time on target into the component times continuously on target. Since counters could not be reset to zero during the short intertrial intervals, i.e., 0, and 15 sec., the apparatus was photographed at the end of each trial with a 16-mm. movie camera.

In all conditions the rotor was allowed to run continuously throughout the session. The *E* timed the trials and the rest intervals with a stop watch and orally gave "ready," "start," and "stop" signals to *S*. The plate circuit of the Tracking-time Analyzer was turned on only during the trials, thereby controlling the time during which *S* might score.

The operation of the Tracking-time Analyzer was clearly audible to all *Ss*, although it and the rotor were separated by a partial wall. The chattering of the relays and the counters provided *S* with limited knowledge of results since the analyzer operated only when *S* was on target. The calibration of the Tracking-time Analyzer was checked twice a day, morning and afternoon, because of changes in the line voltage.

Conditions.—All *Ss* received 21 30-sec. trials. One group of 20 *Ss* received these trials con-

secutively without rest. The other four groups had 15-, 30-, 45-, or 60-sec. intertrial rest. Following Trial 21 all *Ss* had a 5-min. rest. During the 5-min. rest, *Ss* read or talked with *E*. Following this rest, all groups had nine 30-sec. trials under massed practice. The distributed-practice *Ss* were instructed during the 5-min. rest that the conditions would be changed somewhat and that they would have nine consecutive trials without rest.

Instructions.—All *Ss* were given general instructions on the nature of the task, including suggestions and a demonstration of the best way to practice. The *Ss* were told that they would have 21 trials followed by a 5-min. rest and then approximately 5 min. of massed practice. The group which had massed practice (0-sec. rest between trials) throughout was told that its prerest trials would consist of one long 10-min. trial. The groups which had intertrial rests were told that between each of the trials there would be a 15- (30-, 45-, 60-) sec. rest. No questions as to the nature of the scoring apparatus were answered until the session was over.

Results

Prerest Performance

For any effect to be attributed to distribution of practice, all groups must be initially comparable in performance. An analysis of variance was performed on the total-time-on-target scores for Trial 1. The differences among group means were not significant ($F = 1.38$, 4 and 95 *df*, $P > .10$). In addition, a chi-square test was made for differences among the distributions of durations of hits at the first stage of practice, i.e., the first block of three trials. Again the initial differences among groups were not significant ($\chi^2 = 18.27$, 16 *df*, $P > .30$).

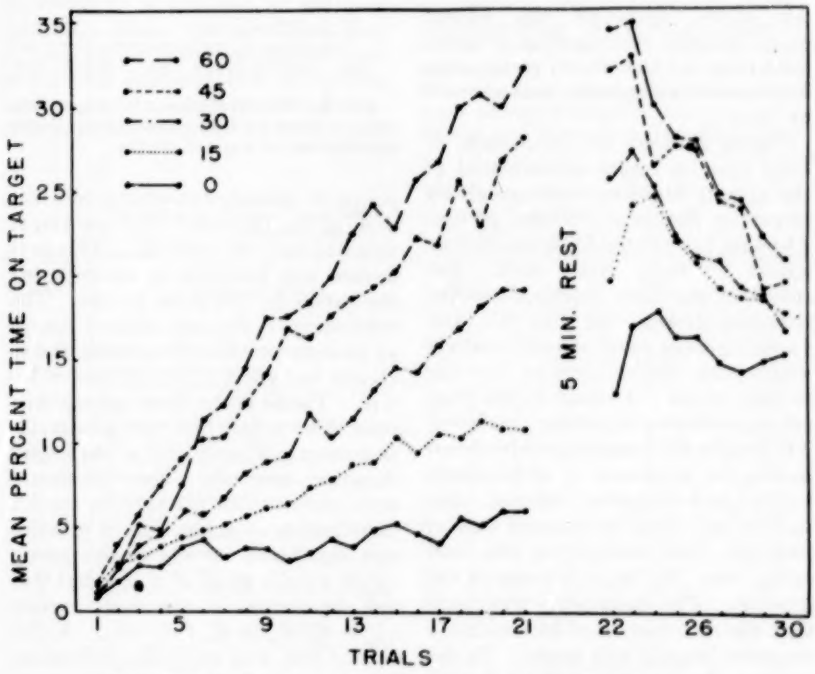

Fig. 1. Mean percentage of total time on target as a function of pre- and postrest practice and intertrial rest intervals.

Figure 1 represents the usual method of presenting rotary pursuit data. The mean percentage of total time on target is plotted as a function of practice. This figure demonstrates graphically the facilitation of performance due to distributed practice. Statistical verification of this facilitation is provided by an analysis of variance of the total time-on-target scores for Trials 19, 20, and 21 combined. Orthogonal polynomials applied to these means show only a significant linear component ($F = 29.28$, 1 and 95 *df*, $P < .01$). Clearly the optimal level of practice distribution was not reached with the values of the independent variable in the present study. Facilitation is evident fairly early in practice and the groups differentiate in performance after the first few trials. Although the 45-sec. group initially performed at a higher level than did the 60-sec. group, that difference was eliminated and reversed by Trial 6.

Figure 2 shows the percentage of total time on target accumulated at the various durations continuously on target for the 0- and 60-sec. groups. The data are plotted for seven prerest blocks of three trials each. The curves of the three intermediate distribution groups, the 15-, 30-, and 45-sec. groups, lay almost entirely within the limits defined by the plotted curves. To simplify the presentation they have not been included. The graphs were constructed by determining the mean number of hits made within each duration interval, correcting this value for unequal sizes of intervals, and multiplying this corrected value by the mid-point of the interval. The assumption was made that the distribution of hits within a duration interval was linear. To determine the *percentage* of total time accumulated at each of the duration intervals, the products of the mid-points × corrected number of hits was divided by the total time on target summed over all intervals. This procedure was followed in constructing the curves for the seven blocks. The separation of the two plotted curves as prerest practice progresses shows clearly the effect of distributed practice. The *S*s in the 60-sec. group shift from short to long hits, thus accumulating more and more time at the longer duration intervals. Two chi-square tests showed that the shift in the hit distribution as a function of practice was significant for the 60-sec. group ($\chi^2 = 442.43$, 48 *df*, $P < .01$) but was not significant in the 0-sec. group ($\chi^2 = 49.60$, 36 *df*, $P > .05$). A chi-square test was made for differences among groups in the continuous-time-on-target distributions on the last block of prerest trials. The chi square

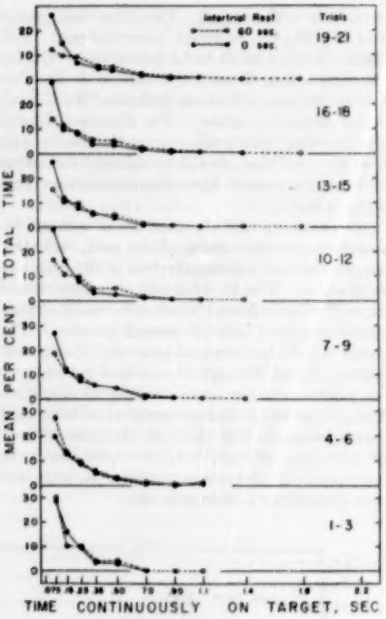

FIG. 2. The distributions of times continuously on target for 0- and 60-sec. groups during the seven prerest stages of practice.

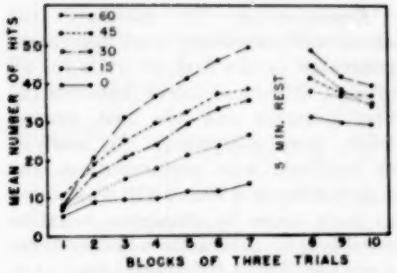

FIG. 3. Mean number of hits, regardless of duration, as a function of pre- and postrest practice and intertrial rest intervals.

was significant at the .01 level ($\chi^2 = 363.14$, 24 *df*) again indicating facilitation of performance as a function of distributed practice.

Figure 3 shows the mean number of hits, for the five groups, regardless of duration, as a function of stage of practice. These curves suggest that at least some improvement in performance is attributable to *S*s learning how to get on target more frequently. It is apparent that the functions tend to be negatively accelerated, especially for the more highly distributed groups.

Postrest Performance

Figure 1 shows the postrest performance of the five groups during the nine massed trials. A modification of the Alexander trend test (**2**) employing orthogonal polynomials was used to test for differences among the postrest time-on-target scores. A summary of this analysis is shown in Table 1. The differences among the group means are significant at the .01 level (Line *a*). The curves bear the same ordinal relation in terms of goodness of performance as they did in the prerest phase of practice. The test shows that the between-group slopes (Line *c*) and between-group cubics (Line *g*) components are also significant sources of variation. The slopes of the curves generally increase in negativity as the prerest intertrial interval increases. The differences between group cubic components suggest the superposition of a higher-order function upon the linear. Figure 1 shows that the 0-, 15-, and 30-sec. groups tend to rise to a maximum postrest performance level before declining, whereas for the 45- and 60-sec. groups this rise is not as marked. The between-*S* slopes (Line *d*) and between-*S* quadratics (Line *f*) as well as the between-*S* means (Line *b*) components indi-

TABLE 1

SUMMARY OF ALEXANDER TREND ANALYSIS OF VARIANCE OF TIME-ON-TARGET SCORES DURING POSTREST TRIALS

Line	Source of Variation	Error Term	*df*	Mean Square	*F*
a.	Between-group means	b	4	367.154	3.742**
b.	Between-individual means	q	95	98.121	65.756**
c.	Between-group slopes	d	4	58.624	12.295**
d.	Between-individual slopes	q	95	4.768	3.195**
e.	Between-group quadratics	f	4	3.819	1.047
f.	Between-individual quadratics	q	95	3.648	2.445**
g.	Between-group cubics	h	4	7.209	3.313*
h.	Between-individual cubics	q	95	2.176	1.458
i.	Between-group quartics	j	4	3.037	1.387
j.	Between-individual quartics	q	95	2.189	1.467
k.	Over-all slope	q	1	650.591	435.994**
l.	Over-all quadratic	q	1	2.617	1.753
m.	Over-all cubic	q	1	15.295	10.250**
n.	Over-all quartic	q	1	11.816	7.919**
o.	Over-all deviations from estimation	q	4	4.636	3.107
p.	Between-group deviations from estimation	q	16	1.980	1.327
q.	Between-individual deviations from estimation		380	1.492	
r.	Total		899		

* $P < .05$.
** $P < .01$.

cate the significant differences among *Ss*. In view of the finding of significant differences among the group slopes, the over-all tests are less meaningful. As would be expected, the over-all slope was significant (Line *k*), indicating that for these nine postrest trials performance decreases. Because significantly different functions have been summed, we would expect to find over-all postrest performance in this experiment to be a complex function of pre- and postrest conditions of practice.

The distributions of times continuously on target during the postrest phase of practice, presented in Fig. 4, show a trend opposite to that observed during the prerest trials. With massed practice, the distributions of the five groups tend to converge. A chi-square test of the distributions on the initial postrest trial (Trial 22) showed that the differences among the groups were significant at the .01 level ($\chi^2 = 149.91$, 24 *df*). Figure 4 demonstrates graphically the near coincidence of the 0- and 60-sec. groups at the end of nine massed practice trials. A chi-square test of the distributions on the last postrest trial (Trial 30) indicated that the groups no longer differed significantly ($\chi^2 = 24.13$, 20 *df*, $P > .20$). An analysis of variance performed on the time-on-target scores for Trial 30 confirmed the convergence of the groups. No significant differences among group means were found ($F = .926$, 4 and 95 *df*, $P > .10$).

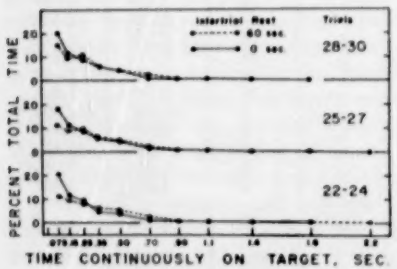

FIG. 4. The distributions of times continuously on target for 0- and 60-sec. groups during the three postrest stages of practice.

Reminiscence.—To estimate the amount of temporary work decrement generated in the first 21 trials for all groups, difference scores, between the first postrest and the last prerest trials, were computed. An analysis of variance was performed on the coded difference scores (10 was added to each score to eliminate negative numbers) to determine whether there was a significant differential amount of reminiscence. The analysis indicated significant differences among group means ($F = 2.62$, 4 and 95 *df*, $P < .05$). Although it appears from the data that the maximum amount of reminiscence occurs in the 15-sec. group, an orthogonal polynomial analysis applied to the group means indicated a significant linear component only ($F = 7.49$, for 1 and 95 *df*, $P < .01$). Apparently, the finding of more reminiscence after a 15-sec. intertrial rest was used was a chance event.

A chi-square test was made for the differences between the frequency distributions for durations continuously on target before and after the 5-min. rest. Table 2 shows these distributions. The shifts in the distributions of the 0- and 15-sec. groups were significant at the .01 level ($\chi^2 = 28.16$ and 19.95, with 3 and 5 *df*, respectively). The chi-square test for the 45-sec. group also proved significant ($\chi^2 = 16.11$, 8 *df*, $P < .05$) although that for the 30- and 60-sec. groups did not ($\chi^2 = 4.66$ and 1.87, with 5 and 8 *df*, $P > .30$ and $P > .98$, respectively). Improvement in performance as a result of the 5-min. rest takes the form of an increase in number and shift in duration of hits.

Warm-up effect.—As a check on the effect of a warm-up decrement, an analysis of variance was performed on the time-on-target scores for Trials 22 and 23. This analysis verified the significance of the initial postrest

TABLE 2

TOTAL NUMBER OF HITS PER DURATION INTERVAL ON LAST PREREST AND FIRST POSTREST TRIALS

Group	Trial	Time Continuously on Target												Sum
		.00	.05	.10	.20	.30	.40	.60	.80	1.00	1.20	1.60	2.00	
60	21	197	281	219	129	50	76	21	8	5	1	1	0	988
	22	214	274	232	238	49	82	24	12	5	2	1	1	1034
45	21	188	194	170	84	51	59	25	10	10	1	1	2	795
	22	152	223	220	87	64	83	19	7	15	2	0	0	877
30	21	199	213	153	80	28	39	3	1	2	3	0	0	721
	22	243	269	191	98	49	59	13	8	2	0	0	0	932
15	21	159	209	101	44	7	13	1	0	1	0	0	0	535
	22	194	266	150	65	36	32	7	1	1	1	0	0	753
0	21	102	101	44	22	4	7	1	0	1	0	0	0	260
	22	150	233	102	52	17	17	6	1	0	0	0	0	578

differences among group means (F = 17.82, 4 and 190 *df*, $P < .01$) but indicated no significant effect due to trials, i.e., the apparent increase in performance from Trial 22 to Trial 23 was not significant (F = 1.26, 1 and 190 *df*, $P > .10$). Warm-up decrement is not evident in the present data. Furthermore, since the Trials × Groups interaction term was not significant, it can also be concluded that the apparent differential increase in performance from the first to the second postrest trials is not a statistically reliable phenomenon.

DISCUSSION

The obtained results may be interpreted as evidence for the existence of a permanent work decrement or conditioned inhibition, as proposed by Ammons (3), Hull (9), and Kimble (10). There are significant differences among the performances of the five groups after a 5-min. rest. However, an alternative to the ${}_{S}I_{R}$ explanation of this difference might be considered. Initially, all *S*s make relatively uncoordinated pursuit movement, i.e., jagged, saw-tooth, circular responses. We propose that there is only one type of inhibition and that this has essentially the same properties as Hull's reactive inhibition or Ammons' temporary work decrement. As in their systems, it develops at a negatively accelerated rate and dissipates spontaneously during rest. However, in addition to reducing performance, we propose that it interferes with the acquisition of skill and thereby reduces the actual level of learning of the habit. In effect our alternative notion argues that the massed-practice *S*s are inferior to the distributed-practice *S*s on the final prerest trials for two reasons: (*a*) their performance is being interfered with, and (*b*) they have learned less of the smooth pursuit tracking skill. The dissipation of inhibition during the 5-min. rest reduces the interference with performance but the level of the skill remains unchanged. Consequently, we believe that the inequality in performances of the groups on the first postrest trials is not due to conditioned inhibition or permanent work decrement but rather to inequality in learning smooth circular tracking. Evidence for this hypothesis comes from the times-

continuously-on-target performance measure. At the outset of practice, nearly all the hits made by *S*s are of short duration, indicating an inability to remain on target. As practice proceeds, the groups differentiate in ability to make longer hits, i.e., to make coordinated circular movements. The intial postrest difference is due primarily to the difference in level of skill which varies in accordance with the length of intertrial interval and the amount of inhibition sustained. Under conditions of massed practice, this difference is eliminated and there is a shift back to less coordinated pursuit with an accompanying increase in short-duration hits. An hypothesis of this general nature would eliminate the need for a construct of permanent work decrement or conditioned inhibition.

The results of this experiment suggest that improvement in performance can be attributed to a combined increase in number and length of hits. However, perfect performance is one hit of 30-sec. duration on any given trial. Since the mean number of hits for the 60-sec. group on Trial 21 is near 50, the total number of hits must decrease as improvement in performance takes place later in learning. The hypothesis of initial increase followed by a decrease as learning progresses is partly substantiated by the negatively accelerated curves of Fig. 4, although even after 21 distributed trials the maximum number of hits has not been attained. In terms of actual performance, this hypothesis suggests that *S* first learns how to get on target, later how to stay on.

An interesting comparison can be made between these data and the data of Archer, Kent, and Mote (7). Using the same Tracking-time Analyzer, these *E*s measured performance in the Pedestal Sight Manipulation Test (PSMT). They found that total number of hits first decreased and then increased, apparently a contradiction of the proposed hypotheais. The PSMT, a form of three-dimennional tracking, presents *S* with a more complex task than the pursuit rotor. These *E*s found the initial decrease in hits attributable to a more rapid decrease in number of short hits compared with the rate of increase in number of longer hits, total time on target remaining relatively stable. The secondary increase in both number and duration of hits then resulted in an increase in time on target. The *S*s performing on the PSMT initially make jerking, saccadic movements and later learned to coordinate their responses in relation to the three-dimensional path of the target. Nearly all hits at the outset of practice were .05 to .10 sec. in duration, whereas on the pursuit rotor, where the target path is predictable, *S*s commonly make hits of .30 sec. in the beginning of practice. A comparison of the results of the present study and the previous PSMT study suggests that the frequency distribution of duration of hits is a function of both stage of practice and degree of task complexity. The results indicate that for simple tasks such as rotary pursuit there will be first an increase and then a decrease in number of hits, whereas for more complex tracking tasks there will be first a decrease and then an increase in number of hits comparable to the initial increase of simpler tasks.

Summary

Five groups of 20 female *S*s participated in a rotary pursuit experiment. All *S*s had thirty 30-sec. trials on a 54-rpm pursuit rotor. Each group worked under a different condition of practice distribution with 0-, 15-, 30-, 45-, and 60-sec. intertrial rest. After 21 trials, each group received 5 min. of rest followed by nine massed-practice trials. Two response measures were used, total time on target and time continuously on target, i.e., duration of hits.

The essential findings were: (*a*) Distribution of practice facilitated performance. (*b*) After the 5-min. rest, a significant amount of reminiscence was shown in the 0- and 15-sec. groups. (*c*) The differences among groups in performance on the first postrest trial were significant. This *may* be taken as evidence for the existence of a permanent work decrement. However, a more parsimonious explanation requiring but a single inhibitory construct was offered. (*d*) Under conditions of postrest massed practice, the groups tend to converge so that by Trial 9 the differences among the groups were not significant. (*e*) Un-

like previous studies, a significant warm-up effect was not observed.

Evidence that improvement in performance takes the form of more hits initially and shifts to longer hits later in practice was presented and discussed. The results of the present experiment indicate that the skill of staying on target for relatively long durations continuously is less well learned by *S*s serving under massed practice as compared to *S*s serving under distributed practice. Through the use of a Tracking-time Analyzer a measure of this skill was obtained.

References

1. Adams, J. A., & Reynolds, B. Effect of shift in distribution of practice conditions following interpolated rest. *J. exp. Psychol.*, 1954, **47**, 32–36.
2. Alexander, H. W. A general test for trend. *Psychol. Bull.*, 1946, **43**, 533–557.
3. Ammons, R. B. Acquisition of motor skills: I. Quantitative analysis and theoretical formulation. *Psychol. Rev.*, 1947, **54**, 263–281.
4. Ammons, R. B. Acquisition of motor skill: II. Rotary pursuit performance with continuous practice before and after a single rest. *J. exp. Psychol.*, 1947, **37**, 393–411.
5. Ammons, R. B. Effect of distribution of practice on rotary pursuit "hits." *J. exp. Psychol.*, 1951, **41**, 17–21.
6. Archer, E. J. Postrest performance in motor learning as a function of prerest degree of distribution of practice. *J. exp. Psychol.*, 1954, **47**, 47–51.
7. Archer, E. J., Kent, G. W., & Mote, F. A. Effect of long-term practice and time on target information feedback on a complex tracking task. *J. exp. Psychol.*, 1956, **51**, in press.
8. Barch, A. M. The relative contributions of I_R and sI_R to the total inhibitory potential. *Perc. & Mot. Skills Res. Exch.*, 1952, **4**, 27–31.
9. Hull, C. L. *Principles of behavior*, New York: D. Appleton-Century, 1943.
10. Kimble, G. A. An experimental test of a two-factor theory of inhibition. *J. exp. Psychol.*, 1949, **39**, 15–23.
11. Kimble, G. A. Performance and reminiscence in motor learning as a function of degree of distribution of practice. *J. exp. Psychol.*, 1949, **39**, 500–510.
12. Kimble, G. A., & Shatel, R. B. The relationship between two kinds of inhibition and amount of practice. *J. exp. Psychol.*, 1952, **44**, 355–359.
13. Schucker, R. E., Stevens, L. B., & Ellis, D. S. A retest for conditioned inhibition in the alphabet printing task. *J. exp. Psychol.*, 1953, **45**, 97–102.
14. Starkweather, J. A., & Duncan, C. P. A test for conditioned inhibition in motor learning. *J. exp. Psychol.*, 1954, **47**, 351–356.
15. Wasserman, H. N. The effect of motivation and amount of pre-rest practice upon inhibitory potential in motor learning. *J. exp. Psychol.*, 1951, **42**, 162–172.

(Received February 7, 1955)

Journal of Experimental Psychology
Vol. 51, No. 1, 1956

STIMULUS SEQUENCE AND CONCEPT LEARNING[1]

MARVIN H. DETAMBEL[2] AND LAWRENCE M. STOLUROW[3]

University of Illinois

This study is concerned with rules for arranging stimulus information to facilitate concept learning. The rules that gain support may not apply to all concept learning situations, but are expected to be applicable to situations in which the following four conditions obtain. These are: (*a*) *S* must learn several concepts; (*b*) each stimulus item is a positive instance of one concept and a negative instance of all alternative concepts; (*c*) every stimulus item is a compound of several variable components, only one of which determines the concept of which the item is a positive instance; and (*d*) the relevant stimulus component of each concept is unique, rather than including several discriminable values that can occur interchangeably in instances of the same concept.

The individual members of a set of stimulus items that meet these requirements can be presented one at a time in a variety of different orders. Under the training procedure used in this experiment, each item, when presented, constitutes a trial on which *S* indicates the concept that he *thinks* is represented, and on which he is informed which concept *is* represented. From one trial to the next both relevant and irrelevant components may have either the same value or a different value.

The relationship between a relevant and an irrelevant component can be said to be *synchronous* if *both maintain* their value, or if *both change* value from one trial to the next. Conversely, when a change in value of a relevant or irrelevant component is accompanied by no value change in the other, this relationship may be designated *asynchronous*. It is possible for synchrony and asynchrony to occur in a single pair of trials if each of the stimulus items consists of several independently variable components.

It is the thesis of this paper that in the class of stimulus situation defined, the training efficiency of a series of stimulus items is negatively related to the relative degree or frequency of synchrony in the series.

To illustrate the way in which increased synchrony of relevant and irrelevant stimulus components results in decreased information about the concepts, consider two successive trials of a concept learning experiment. In this hypothetical situation the stimulus event on each trial is made up of three components (a stimulus triad). The *S* has been told that from trial to trial each component can assume a different value and that only *one* of these components is relevant to the response he is to make. It is *S*'s task to discover which component is relevant and determine what values of

[1] This research was supported in part by the United States Air Force under Contract No. AF 33(038)-25726 monitored by AFPTRC. Permission is granted for reproduction, translation, publication, use, and disposal in whole and in part by or for the United States Government. The authors are grateful to Drs. W. M. Gilbert, J. F. Kamman, Mr. C. H. Twitty, and personnel of the Student Counseling Bureau, University of Illinois, for their assistance.

This paper was presented in part at the Annual meeting of the Midwestern Psychological Association, April 1955.

[2] Now at The RAND Corporation, Santa Monica, California.

[3] Now at Training Aids Research Laboratory, Air Force Personnel and Training Research Center, Chanute Air Force Base.

that component are associated with each response. Now consider the maximum information that can be extracted by different *S*s from pairs of trials under three sequences that differ in synchrony. On the first trial of the task the same stimulus triad is used for all groups; it is three adjacent rectangles which are red, white, and blue, respectively, as seen from left to right. After seeing this item, all *S*s are told that response "A" is correct. Now consider the second trial. In the first of these sequences the *S*s may see a triad which is blue, white, blue and be told that the correct response is "B." Therefore the information *S*s could derive from this pair of trials would be that the left component is relevant and that the color red meant response "A," and blue "B."

Now consider the second trial of another sequence. This time the second triad might be blue, blue, blue and the correct response "B." The *S*s given this pair of trials could extract less information than those given the first sequence. This second group could eliminate the component on the right since it stayed blue when the required response changed from "A" to "B." However, since the left and center rectangles both changed color, either one could be relevant and at least one additional trial would be needed before *S*s could determine whether the left or the center rectangle was relevant.

In a third sequence, all three components could change value on the second trial. In this case *S*s would be unable to eliminate even one of them from further consideration.

These three examples illustrated different degrees of synchrony, and by applying our experimental hypothesis we should expect to find that the first procedure was best for concept learning and the third one least facilitative.

The principles of stimulus sequencing used in the experiment to be reported are those described in this illustrative case. However, *S*'s task was more difficult in that they were given less preliminary information about the nature and number of relevant components.

Method

One set of 27 projected drawings was used. The sequence in which the 27 stimulus triads were presented during training was different for each group. This was done to produce different degrees of synchrony.

Stimulus Material

Each triad consisted of three adjacent rectangles, all of which had their base at the same vertical level. Each rectangle could have any one of three heights. In the drawings photo-

TABLE 1

Sequence of Stimulus Triads for the Two Training Procedures and the Test That Followed Training*

Trial	Training Procedure						Test Sequence		
	I (Min.)			II (Max.)					
	Rectangle			Rectangle			Rectangle		
	1	2	3	1	2	3	1	2	3
1	H	H	H	H	H	H	M	L	L
2	M	H	H	M	L	L	L	M	H
3	L	H	H	L	M	M	H	L	H
4	L	M	L	L	M	H	M	H	M
5	M	M	L	M	H	M	L	H	M
6	H	M	L	H	L	L	H	L	M
7	H	L	H	H	L	M	M	M	H
8	M	L	H	M	M	H	M	H	L
9	L	L	H	L	H	L	H	L	L
10	L	M	M	L	H	M	H	H	L
11	H	M	M	H	L	H	L	H	H
12	M	M	M	M	M	L	H	H	M
13	M	H	L	M	M	M	L	L	H
14	H	H	L	H	H	L	M	H	H
15	L	H	L	L	L	H	M	L	M
16	L	L	M	L	L	M	H	H	H
17	H	L	M	H	M	H	M	L	H
18	M	L	M	M	H	L	H	M	L
19	M	M	H	M	H	H	M	M	M
20	L	M	H	L	L	L	M	M	L
21	H	M	H	H	M	M	L	M	L
22	H	L	L	H	M	L	L	H	L
23	L	L	L	L	H	H	H	M	M
24	M	L	L	M	L	M	L	L	L
25	M	H	M	M	L	H	H	M	H
26	L	H	M	L	M	L	L	L	M
27	H	H	M	H	H	M	L	M	M

*The letters "H," "L," and "M" designate the height of the rectangles which comprise the stimulus, and may be interpreted as "high," "low," and "medium." The height of Rectangle 1 determines which letter-naming response is correct; for high, medium, and low rectangles, respectively, the response "A," "B," and "C" are correct; this sequence is the same for both training procedures.

graphed to produce film strip, these heights were 2, 1, and ½ in.; every rectangle was 1 in. wide.

Procedure

Each of the 27 stimulus triads was used only once in both training and test trials for each group.

Training sequences.—Without telling *S*s, the rectangle on the left was designated the *relevant* one. Whenever the relevant rectangle was tall, the appropriate response was "A"; when medium sized, "B"; when short, "C."

In Table 1, Rectangle 1 is the relevant rectangle; it can be seen here that the sequence of values of the relevant stimulus component is the *same* for both training procedures, but that the sequence of values of the irrelevant components —Rectangles 2 and 3—is *different* for the procedures.

The stimulus sequence for the *S*s given Procedure I is a sequence of minimal synchrony of relevant and irrelevant components. It will be noted that in Procedure I, on every pair of trials in which the relevant rectangle changes size, both irrelevant rectangles remain constant in size and whenever the relevant rectangle remains constant, both irrelevant rectangles change in size. Of the two sequences used in the experiment, this one is expected to provide the *better* training.

The stimulus sequence for Procedure II represents *maximum* synchrony, the sequence expected to result in poor performance. Rectangle 2 changes size whenever the relevant rectangle does. Due to the requirement that each of the 27 triads should be used only once, it is not possible also to have both irrelevant rectangles remain constant in size when the relevant rectangle is kept constant. Therefore there are eight pairs of trials in which Rectangle 3 changes size while the relevant rectangle does not.

Subjects.—The *S*s were newly enrolled students at the University of Illinois. Procedure I *S*s were 335 students entering the College of Commerce; Procedure II, 275 students entering the College of Agriculture. While these students were assembled for vocational guidance testing, the experiment was conducted as though it were a part of the testing program. At the conclusion of the experiment the students were told that they had participated in an experiment and that their test scores on the concept learning task would not be used in vocational counseling.

Instructions.—The *S*s were told that the test they were about to take was to find out how quickly and accurately they could make use of information given to them in the test items. The following specific instructions were read:

"On this screen I'll show you a series of pictures of three boxes.[4] Each box can have any one of three heights. It can be a tall box like this (*E* pointed to it), or a short one like this, or a middle-sized one like this.

"Each drawing has a letter-name which is related in some way to the height of the figure. Some drawings have the letter-name 'A,' others 'B,' and others 'C.' Your problem is to learn the way in which the height of the figure determines its name, so that you can correctly name any picture that is shown to you. It may be the height of one of the boxes that tells you what the name is, or the height of two of the boxes, or the height of all three. That is something you will have to find out as you look at the pictures.

"This is the way we'll do it. First, you will see a picture of three boxes. While the picture is there, carefully notice the height of each box.

"Next, the word 'write' will appear on the screen to remind you to write the name of the figure on your answer sheet. [*E* showed 'write' frame.] Decide whether the drawing is an 'A,' 'B,' or 'C,' and write that letter after number '1' in the first column of your answer sheet, and so on.

"After you have had a short interval in which to write 'A,' 'B,' or 'C,' the figure will be repeated and the correct letter name will be indicated to the right of the figure, where 'X' appears on this picture. [Showed example.] In addition to telling you whether your answer was right or wrong, this provides an opportunity for studying the relationship between the letter-name and the height of the figure. Study it carefully, but do not change what you have written.

"After an interval without a picture, you will be shown the next picture. Again you are to decide whether the figure is an 'A,' 'B,' or 'C,' write your decision in the next space on your answer sheet, see what the correct answer is, and then wait for the next picture.

"At first your answers will be guesses, but as you see more and more examples you should learn to recognize the three kinds of figures and write the correct letter every time.

"Before beginning the test I'll run the projector a short time to show you how the boxes will look, and how fast you must write your answers [two illustrative trials]. Is there any question about what you are to do?

"Remember, you are to learn to identify each figure as 'A,' 'B,' or 'C.' While the word 'write' is on the screen, write your answer in the first column of your answer sheet, after the correct number. Do not change any answers."

[4] The word "box" was used throughout the instructions to the trainee in preferences to "rectangle."

Questions were answered by repeating or paraphrasing parts of the instructions.

After 27 training trials the projector was halted and the following instructions read:

"Now you will see the pictures in a different order and you are again to identify each figure as an 'A,' 'B,' or 'C.' You will see the same figures, and the same answers will be correct, but this time you will not be told what answer you should have written. Write your answers in the second column and try to answer correctly as often as possible."

As the instructions indicate, each training film contained four frames per trial: (*a*) a picture of three rectangles; (*b*) the word "write"; (*c*) another picture of the same rectangles plus the correct letter name ("X" was substituted for the correct letter in illustrative trials); and (*d*) a blank frame. Since the Dunning Animatic 16-mm. projector was adjusted to change frames automatically every 7.5 sec., each training trial lasted 30 sec.

During the test series there were only two frames per trial: (*a*) a picture of three rectangles, and (*b*) the word "write." Test trials lasted 15 sec.

During both training and test series, *E* announced the trial number about 1 sec. before the stimulus appeared.

Procedure I data were obtained in the morning, and Procedure II data in the afternoon of the same day so that any intergroup communication would tend to improve the performance of those trained under Procedure II and work against the experimental hypothesis.

Results

Since the ACE scores were available on all but 29 *S*s, norms developed on previous University of Illinois students were used to divide *S*s trained under each procedure into five ACE groups. Having determined group membership, the hypothesis was tested that the proportion of solvers was the same for the minimum and maximum synchrony procedures. By making independent checks of the hypothesis at each ACE level it was possible to determine the effectiveness of synchrony regardless of variations in the size of the groups at the different levels for the two procedures and at the same time maintain approximate equivalence in scholastic aptitude. Fifteen such checks were made, using groups selected on the basis of (*a*) Q or quantitative score, (*b*) L or language score and (*c*) total score.

Table 2 shows the number and percentage of *S*s in each ACE group who correctly identified all 27 test items. The percentages are printed in bold face for all ACE quintiles at which the high and low synchrony sequences produced significantly different percentages of solvers. Seven chi squares exceeded the minimum confidence level of .01, and all seven were obtained in the first three quintiles. In no case was there a significant difference between the proportions of solvers at the highest quintiles under the two sequences. These data demonstrate that at low ACE levels, the maximum synchrony sequences produced a significantly smaller proportion of solvers than did the minimum synchrony sequence, but that at high ACE levels the two degrees of synchrony did not result in significantly different proportions of solvers. Since each chi square was computed at an ACE quintile level, it was independent of those at the other quintile levels. Therefore, it was possible to obtain an over-all estimate of the effects of synchrony by summing the chi-square values. The chi squares for the Q score, L score and Total score, with 5 *df* (one per ACE level), exceeded the minimum confidence level of .01, viz. 21.93, 32.60, and 35.36, respectively. The mean percentage of solvers for each sequence can be estimated for the case where there are equal *N*s in the five ACE groups run under a simple synchrony procedure by giving equal weights to every group at each level. This was accomplished by computing the mean of the means in Table 2; those for the minimum synchrony sequence vary from 64.41 to 64.91 and for the maxi-

TABLE 2

NUMBER AND PERCENTAGE OF "SOLVERS" AT EACH ACE QUINTILE FOR THE MINIMUM AND MAXIMUM SYNCHRONY PROCEDURES

Quintile (U. of I. Norms)	Q-Score						L-Score						Total Score					
	Minimum Synchrony			Maximum Synchrony			Minimum Synchrony			Maximum Synchrony			Minimum Synchrony			Maximum Synchrony		
	N	%	Rank of %	*N*	%	Rank of %	*N*	%	Rank of %	*N*	%	Rank of %	*N*	%	Rank of %	*N*	%	Rank of %
1	38	60.32	2	13	24.53	1	50	66.67	3	35	38.56	2	47	64.58	3	27	31.03	1
2	47	66.20	4	24	32.00	2	46	61.33	2	24	35.29	1	35	57.14	1	29	42.65	3
3	29	59.18	1	22	46.81	3	48	70.59	5	20	42.55	3	51	72.86	5	20	38.46	2
4	47	70.15	5	32	54.24	4	33	56.90	1	23	53.49	4	39	63.93	2	21	58.33	4
5	45	66.18	3	19	65.52	5	29	69.05	4	8	57.14	5	33	64.71	4	13	65.00	5
Total	206			110			206			110			206			110		
Mean		64.41			44.62			64.91			45.39			64.60			47.09	

Note.—Bold face used where differences between minimum and maximum synchrony in percentage of solvers are significant beyond the .01 level.

mum synchrony sequence from 44.62 to 47.09, leaving little doubt that the two sequences had different effects.

A further indication of the relationship between scholastic aptitude and performance in this concept learning study can be obtained from Table 2 by comparing (a) the rank orders of the percentages of solvers under each procedure with (b) the ACE quintile numbers. It can be seen that under the maximum synchrony sequence, the percentage of solvers increased regularly with increase in ACE Q score. This perfect correspondence of ranks is significant at the .01 confidence level. Similar comparisons of ranks for percentages of solvers and quintiles (using both the ACE L-score and Total-score quintiles) show one inversion of ranks in each case. These rank correlations are significant at the .05 level. In the case of minimum synchrony, the sequence expected to result in better learning, there was no significant correlation between the percentage of solvers and any of the three sets of ACE quintiles.

In a final analysis using data from Table 2, no significant differences were found when the percentage of solvers in the Quintile 5, maximum synchrony group, was compared with the percentage of solvers in each of the quintiles under the minimum synchrony procedure.

The data summarized in Table 2 indicate an interaction between scholastic aptitude and amount of synchrony. Under the minimum synchrony procedure, *S*s at all ACE levels were about equally successful in learning the concepts. Under the maximum synchrony procedure, however, there was a positive correlation between scholastic aptitude and solution of the concept learning problem. At low ACE levels there were significantly fewer solvers under this procedure than under the minimum synchrony procedure, but maximum synchrony *S*s at the top ACE levels were about as successful in learning the concepts at were the minimum synchrony *S*s as all levels.

The interaction just described indicates that the ACE level of groups trained under difference degrees of synchrony is an important determiner of the amount by which those groups will differ in percentage of solvers.

For example, the data in Table 2 suggest that if only *Ss* with high ACE scores had been used in this experiment there would have been no significant difference between maximum and minimum synchrony in percentage of solvers. On the other hand, this difference would have been maximized if only *Ss* with low ACE scores had been used.

The foregoing analyses of performance on the concept learning task were concerned only with the proportions of *Ss* who obtained perfect or less than perfect scores on the test series. Two more detailed performance summaries are the frequency distributions of training scores shown in Fig. 1 and the distributions of test scores in Fig. 2. In testing the hypotheses that the frequency distributions obtained under the two degrees of synchrony are random samples from the same populations the "0–3" and "4–7" score categories were not used because of small theoretical frequencies. The χ^2 for the training score distribution was 79.4 and for the test distribution 31.3. Both results are statistically significant beyond the .01 level (with 4 *df*, $\chi^2 = 13.277$ has $P < .01$), and we therefore conclude that *Ss* trained under the two procedures differed in performance during both training and test series, with *Ss* in the minimum synchrony procedure tending to respond correctly more frequently than *Ss* in the maximum synchrony procedure.

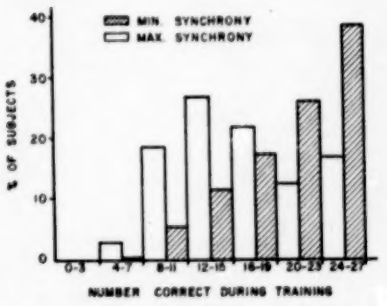

FIG. 1. Training score distributions for Procedure I (minimum synchrony) and Procedure II (maximum synchrony).

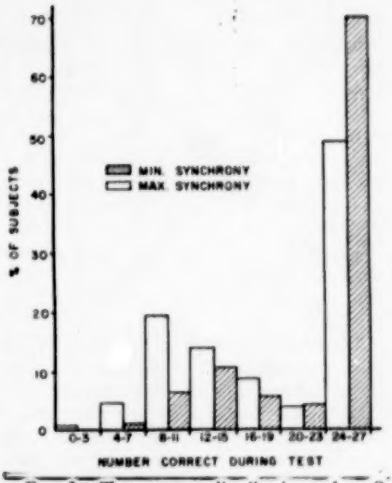

FIG. 2. Test score distributions for *Ss* trained on Procedure I (minimum synchrony) and for *Ss* trained on Procedure II (maximum synchrony).

DISCUSSION

A number of previous studies have been concerned with the relative effectiveness of positive and negative instances in concept learning (1). The present study points out that for a given set of positive and negative instances, the *sequence* in which they are presented is an important factor in training effectiveness. Specifically, the results strongly support the experimental hypothesis that, in the class of concept learning situations under consideration, the training efficiency of a sequence of items is negatively related to the degree of synchrony of relevant and irrelevant stimulus components occurring in the sequence.

For the experimenter in concept learning this result suggests that synchrony is a variable to be controlled in future experiments.

For the person interested in efficient training procedures this result suggests rules for arranging stimulus sequences. If one is to learn a set of concepts by observing a sequence of instances of them, then: (*a*) When the same relevant cue is presented on adjacent trials, as many as possible of the irrelevant stimulus components should have different values on the two trials. (*b*) When two different relevant cues are presented on adjacent trials, as many as possible of the irrelevant stimulus components should be kept constant.

The reader is reminded again that these rules may not apply to all concept learning situations, but are expected to apply in at least those situations which meet the conditions specified in the first paragraph of the paper. Furthermore, these rules are not intended as a complete set even for situations which meet these conditions, for other rules of efficient procedure have yet to be determined. Missing, for example, are the rules for generating an optimum sequence of relevant cues.

The greater effectiveness of the low synchrony sequence appears to derive at least in part from an interesting and readily understandable sort of interaction between scholastic aptitude and degree of synchrony. At low synchrony levels scholastic aptitude and concept learning have no significant correlation, while at high synchrony levels there is a significant positive correlation between these variables. Apparently, *S*s at all aptitude levels do equally well in learning the concepts as long as the stimulus sequence contains little synchrony of relevant and irrelevant stimulus components. However, as synchrony is increased, fewer and fewer of the low aptitude *S*s learn the concepts.

Summary

An experiment was conducted to investigate relationships between concept learning and the degree of synchrony of relevant and irrelevant stimulus components that occurs during the learning period. Two stimulus components are said to be in synchrony if both maintain their value, or both change value from one trial to the next.

Every stimulus item was a compound of three variable components (a stimulus triad), only one of which was relevant to the response *S* was to make. Each stimulus component was a rectangle that could have any of three values (i.e., any of three heights), and for each value of the relevant component a different letter of the alphabet was the correct response.

A single set of 27 stimulus triads was used, but the sequence in which the triads was shown during training differed in the two procedures. However, while the sequence of stimulus triads differed, the sequence of relevant cues and therefore the sequence of correct responses was identical in both treatments. Thus, the two training procedures differed in the degree of synchrony of relevant and irrelevant components.

An interaction was found between the degree of synchrony in the stimulus sequence and scholastic aptitude as measured by the ACE examination. For *S*s with high ACE scores the two stimulus sequences did not produce significantly different proportions of *S*s who learned the concepts. However, as ACE score decreased the proportion of "solvers" under the high synchrony procedure decreased, becoming significantly smaller than the proportion of "solvers" under the low synchrony procedure. The proportion of "solvers" under the low synchrony procedure did not change significantly as a function of ACE score. Consistent with this interaction, the proportion of "solvers" was significantly greater under the low synchrony sequence than under the high synchrony sequence. These results supported the experimental hypothesis that, in the class of situations considered, the training efficiency of a series of stimulus items is negatively related to the degree of synchrony of relevant and irrelevant stimulus components in the series.

Reference

1. Hovland, C. I. A "communication analysis" of concept learning. *Psychol. Rev.*, 1952, **59**, 461–472.

(Received November 13, 1954)

Journal of Experimental Psychology
Vol. 51, No. 1, 1956

PROBABILITIES OF SUCCESS AND AMOUNTS OF WORK IN A MULTICHOICE SITUATION [1]

MARVIN H. DETAMBEL

University of Southern California [2]

This experiment was performed to gain some information about the relative weighting of two variables in repetitive multichoice situations. The two variables are (*a*) the relative frequencies with which the several possible choices are "correct," and (*b*) the amounts of work associated with the alternative choices.

These variables are prominent in the experience of men who diagnose and repair malfunctioning equipment. Frequently a trouble symptom indicates only that the trouble, or some additional information about the trouble, lies in one of several possible locations. Given such information, the trouble shooter must decide which possibilities to check first, second, etc., until the trouble is identified. These are important decisions, for in the typical case the alternatives differ in both (*a*) their probability of being the actual trouble, and (*b*) the amount of time required to perform the checking operations; under these circumstances there are usually some sequences of checking the alternatives which are more efficient than others, in that they consume a smaller average amount of time in identifying the trouble.

Given information about the probability and work time associated with each of the possible troubles, the relative efficiency of all behavior routes can be determined. For example, suppose a symptom indicates that the trouble lies in component A, B, or C, that these components are the cause of the symptom with relative frequencies .7, .2, and .1, respectively, and that 50 min. are needed to check A, 5 min. to check B, and 4 min. to check C. The probability information, considered separately, suggests checking in A-B-C order, since A is the most likely trouble source and C least likely. The work times suggest checking in C-B-A order since C requires the least time and A the most. However, calculation of the average amount of time per discovery of the trouble via each of the six possible behavior routes indicates that the B-C-A sequence is most efficient.

To illustrate such calculations, consider 100 occurrences of the same symptom and consistent use of the A-B-C checking order. The checking operation will require 50 min. on each of the 70 occasions when A is the trouble source, or a total of $50 \times 70 = 3500$ min. On the 20 occasions when A is not faulty but B is, each check will require $50 + 5 = 55$ min., or a total of $55 \times 20 = 1100$ min. On the 10 occasions when C is faulty, each check will require $50 + 5 + 4 = 59$ min., or a total of $59 \times 10 = 590$ min. Total time for 100 occurrences of the symptom is 5190 min., an average of 51.9 min. per occurrence of the symptom.

Table 1 summarizes a set of similar computations for the six possible checking sequences and shows that the B-C-A order requires the least average amount of time per trouble occurrence. Other sequences require an average of .3 to 9 additional minutes, or up to 20.8% more time per trouble occurrence.

It is of considerable theoretical interest and practical importance to determine whether persons, as a result of repeated experience with probability-work time situations of the sort just described, learn to follow the most efficient behavior route. Since it was hypothesized that in many cases men do not adopt the most efficient

[1] This report covers work performed under the University of Southern California–U. S. Navy Electronics Laboratory Joint Project No. 8. It is part of a series of investigations under Contract No. NObsr-57435 between the Bureau of Ships, Navy Department, and the University of Southern California in cooperation with the U. S. Navy Electronics Laboratory. The experimental work was carried on at the U. S. Navy Electronics Laboratory. Miss Mary Jane Haugen assisted in collecting the data.

[2] Now at The RAND Corporation, Santa Monica, California.

TABLE 1

RELATIVE EFFICIENCY OF VARIOUS CHECKING SEQUENCES WHEN SEVERAL POSSIBLE TROUBLES DIFFER IN PROBABILITY AND CHECKING TIME

Trouble	Probability	Checking Time (min.)	Checking Sequence	Mean Checking Time per Trouble Occurrence* (min.)
A	.7	50	ABC	51.9
B	.2	5	ACB	52.2
C	.1	4	BAC	45.4
			BCA	43.2
			CAB	50.0
			CBA	43.5

* Computational procedure is described in text.

behavior sequence, but check first either the alternative with greatest probability of success or least work, it was decided to find out which of these variables carries greater weight in determining behavior sequences.

To accomplish this, *S*s were exposed repeatedly to a situation in which four alternatives had different probabilities of being the source of trouble and required different work times, but in which the average work time was equal for all of the 24 possible checking sequences. Since no sequence was more efficient than any other, the development of a preferred behavior route could not be attributed to its relative efficiency. It should be fairly straightforward, therefore, to determine whether *S*s adopt a behavior route based primarily upon the probabilities or upon the amounts of work associated with the four alternatives.

METHOD

The simulated trouble-shooting situation in which *S*s worked is described in the following instructions which were read to each *S* at the beginning of the experiment.

"Imagine a piece of equipment that is not functioning properly. You have been trouble-shooting the equipment and you know that the trouble lies in one of four possible locations. Your next job is to check these four possibilities to determine which one is actually causing the difficulty. You want to complete this work in the least possible time, so the order in which you check the four parts may be important. The checking order *will* be important if the checking time is not the same for all four parts, and if the parts are not equally likely to be the cause of the trouble.

"Now, let these four decks of cards represent the four components in which the trouble might lie. You can check a component by turning over one card at a time, like this, until you turn up an information card. There are two types of information cards, the ace of spades and the joker. Whenever you turn up a joker you will know the component is okay and you should begin to check a different deck. The ace of spades represents the faulty component, so whenever you turn it face up you can stop sorting cards and wait for the next trial. On each trial one deck will contain the ace, the others will contain jokers. The ace of spades will not be in the same deck on every trial, but it will appear in some decks more often than others. You will also find that the depth at which the information card is buried is not the same for all decks.

"Each trial will be completed when you have exposed the ace of spades. In looking for the ace try to check the decks in the order which, on the average, requires the smallest amount of time and effort; that is, the order which minimizes the number of cards you must turn to find the ace. As the experiment progresses I will be interested in any comments you may have concerning the most efficient order for checking the four decks. You will be given an opportunity to comment after each trial. Are there any questions?

"Remember, your task is to find the ace of spades in the most efficient manner. Select a deck. Turn over one card at a time until you expose a joker or ace of spades. If a joker, go to another deck and continue to turn cards. Stop only when you have exposed the ace of spades."

The random assignment of the ace of spades to the four decks was such that in every block of 15 trials the ace was found in Deck A eight times, in B four times, in C twice, and in D once. These ace-in-deck frequencies suggest that the efficient order for checking the decks is A-B-C-D. However, in Deck A the information card was always the sixteenth card turned; in B the information card was eighth; in C, fourth; and in D, second. This information about amount of work, when considered by itself, suggests that the decks should be checked in D-C-B-A order.

These particular values of the two variables were selected so that no checking order would be more efficient than any other; for each of the 24 possible orders of checking the four decks the average number of cards turned in finding the

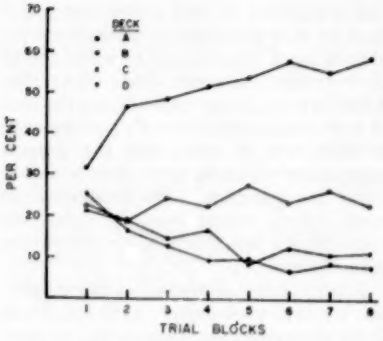

FIG. 1. Percentage of first choices on the indicated deck was selected.

ace is 20.67. Therefore the development of a preference for a particular order could not be attributed to the relative efficiency of that order.

Twelve men served as *S*s, each finding the ace 120 times.

On each trial the decks were presented in four canasta baskets placed in a row on a wooden tray. Two such trays were used so that while *S* was working with one set of four decks an assistant could prepare the other set for the next trial. For any *S*, the location of Decks A, B, C, and D remained constant from trial to trial, but a different arrangement of decks was used for each *S*. Of the 24 possible ways of assigning the four decks to the four baskets, 12 were used, with all decks being assigned equally often to each basket.

RESULTS

Figure 1 shows, as a function of trials, the relative frequency with which each of the decks was the first to be checked. At the outset the four decks were about equally likely first choices. These percentages gradually diverged, with Deck A becoming the most popular first choice, B the second most popular, C third, and D the least popular. The consistency of this preference can be seen in Table 2, which shows each *S*'s distribution of first choices for the final 30 trials.

On some trials the first deck examined did not contain the ace, making it necessary for *S* to select a second deck for examination. Of the three

TABLE 2

FREQUENCY DISTRIBUTION OF EACH S'S FIRST CHOICES DURING THE FINAL 30 TRIALS

S	Deck			
	A	B	C	D
1	30	0	0	0
2	27	2	1	0
3	21	8	1	0
4	18	8	3	1
5	17	4	4	5
6	16	11	2	1
7	16	8	2	4
8	15	9	4	2
9	15	8	4	3
10	12	8	4	6
11	11	11	6	2
12	6	11	8	5
Mean	17.0	7.3	3.2	2.4

decks among which the second choice was made one had the highest probability of containing the ace and required the greatest amount of work, one had the lowest probability and required the least work, and one had an intermediate probability and required an intermediate amount of work. Figure 2 shows, as a function of trials, the relative frequencies with which the second deck checked was the one with highest probability, the one with second highest probability,

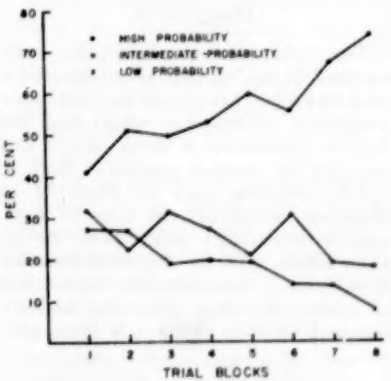

FIG. 2. Percentage of second choices on which the indicated deck was selected.

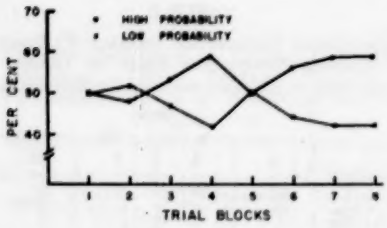

FIG. 3. Percentage of third choices on which the indicated deck was selected.

and the one with lowest probability. Again the percentages diverged, with the deck most likely to contain the ace becoming the most popular choice, and the deck least likely to contain the ace becoming least popular choice.

When neither the first nor second deck contained the ace, a choice had to be made between the two remaining decks. Figure 3 shows, as a function of trials, the relative frequencies with which the high-probability and low-probability decks were selected. As with the first and second decks selected, there is an increasing tendency, with trials, to choose the deck which required more work but was more likely to contain the ace. However, the preference is less marked in this third choice.

DISCUSSION

The *Ss*' performance can be summarized by saying that they acquired a preference for that one of the previously unchecked alternatives which had the highest probability of being correct and required the greatest amount of work.

This outcome may be expected in situations where (*a*) the ratio of work time to probability is the same for all alternatives, and (*b*) the work time for each choice is determined by the number of cycles of some repetitive activity required by that choice. A more general statement of this latter restriction may be that the energy requirement per unit of work time should be equal for all alternatives. It seems quite likely that differences in energy expenditure per unit of work time would alter *S*'s performance in this sort of task, but the present experiment yields no data upon which to argue the question. The interaction of work times, energy requirements, and probabilities remains a topic for future investigation.

Finally, because of the heavier weighting of the probability variable, it is likely that small increases in the amount of work associated with the high probability alternatives would produce little change in *Ss*' behavior; i.e., *Ss* would continue to prefer the high probability-high work alternative even though it has become more efficient to select the low probability-low work alternative. The existence and extent of such inefficient behavior also remains a subject for future research.

SUMMARY

This experiment grew out of a common trouble-shooting situation in which a trouble symptom has a probabilistic relationship to several possible troubles and in which different amounts of work are required in checking the several possibilities. The situation was simulated by repeatedly presenting *S* with four decks of cards, one containing a card representing the trouble, and the others containing cards representing the absence of trouble. The decks differed both in their probability of containing the "trouble" and in the number of cards that had to be turned to discover whether or not the "trouble" was present. However, these probabilities and amounts of work were such that all of the 24 possible sequences for checking the four decks resulted in the same average amount of work per discovery of the "trouble," i.e., for all sequences the same mean number of cards had to be turned per discovery of the trouble indicator. Under these circumstances *Ss* showed an increasing tendency, with trials, to select that one of the previously unchecked decks which had the greatest probability of containing the trouble indicator, even though this choice also required the most work.

Journal of Experimental Psychology
Vol. 51, No. 1, 1956

ACQUISITION, EXTINCTION, AND SPONTANEOUS RECOVERY AS A FUNCTION OF PERCENTAGE OF REINFORCEMENT AND INTERTRIAL INTERVALS [1]

DONALD J. LEWIS

Northwestern University

Sheffield (**19**) and Stanley (**22**) have reported evidence to indicate that distributed (15-min. intertrial intervals) extinction trials result in faster extinction, as determined by running time, than massed (15-sec. interval) trials. These two reports differ from seven studies (**6, 7, 12, 15, 16, 17, 24**) reporting massed extinction trials to result in quicker extinction, and two studies (**13, 14**) which show no difference in effect of intertrial interval.

The effect of the intertrial interval upon acquisition is perhaps more consistent, at least to the extent that there are no studies reporting exactly contradictory evidence. Five recent studies (**3, 8, 16, 21, 24**) show spaced acquisition to result in superior conditioning to massed, and two studies (**6, 18**) show no difference as a result of spacing. No study has been reported showing massed trials to be superior to spaced.

The reports of the interaction effects between spaced and massed trials in acquisition and extinction are also not entirely consistent. After massed acquisition, three studies (**6, 19, 22**) show that spaced trials lead to faster extinction; and one (**16**) shows no difference between spaced and massed. After spaced acquisition, two studies (**16, 6**) show massed trials leading to quicker extinction, one (**19**) shows spaced leading to quicker extinction, although the difference was not quite statistically significant, and one (**22**) shows no difference.

Another variable with important effects for both acquisition and extinction is the percentage of reinforcement administered during acquisition. Many studies (**10**) indicate that a partially reinforced response is acquired more slowly than a continuously reinforced one. Sheffield (**18**), however, found no difference between partial and continuous reinforcement in this respect. The evidence of the effect of partial reinforcement on extinction is quite consistent in showing a partially reinforced response to extinguish slower than a continuously reinforced one. Sheffield (**18**) has presented evidence to show that this "Humphreys effect" is due, in part at least, to the acquisition intertrial interval. Grant, Schipper, and Ross (**6**), and Weinstock (**25**), however, found the Humphreys effect to be independent of the acquisition interval.

The evidence on spontaneous recovery (**4, 5, 7, 11, 26**) although meager, is relatively consistent in showing spontaneous recovery to be a function of time. Extremely few experiments have attempted to determine the effect of other stimulus variables upon spontaneous recovery.

The purpose of the present experi-

[1] The author wishes to express his thanks to Nan A. Lewis for her help in running *Ss* and to Carl P. Duncan for his advice during the statistical analysis and the writing of the report, and to John Cotton for a critical reading of the manuscript. Gratitude is also due to the College of Liberal Arts of Northwestern University for a Summer Research Grant-in-Aid.

ment was to contribute empirical evidence on the action and interaction of: (*a*) two intertrial intervals in acquisition, (*b*) two percentages of reinforcement, (*c*) two intertrial intervals in extinction, and (*d*) four postextinction time intervals prior to a spontaneous recovery test.

METHOD

Subjects.—The *S*s were 80 experimentally naive male albino rats purchased from the Sprague-Dawley Co., of Madison, Wisconsin. Nine animals were discarded during the course of experimentation due to sickness because of the effects of winter temperature variability. These animals were replaced. Age of the animals varied from 75–90 days, with the mean at approximately 80 days, and their mean weight was approximately 200 gm. All *S*s were randomly assigned to the various experimental conditions.

Apparatus.—The dimensions and coloring of the runway used in this experiment conformed to those reported by Sheffield (18). The lighting in the room came solely from an overhead 60-w. bulb. An automatic timer was activated when the starting-box door was raised, and was stopped when an animal put weight on the floor of the goal box.

Procedure.—At least seven days prior to experimentation *S*s were placed on a 24-hr. restricted diet of approximately 7½ gm. of Purina Dog Chow. Each day during these seven, *S*s were handled and tamed for approximately 5 min. The day before their first pretraining trials each *S* was placed for approximately 2 min. in the closed starting box. Then *S* was carried to the closed goal box where he remained until he had eaten three pellets from the food dish, after which he was returned to his living cage. Aside from the time devoted to taming and experimentation, *S* remained in his living cage.

On the day following habituation, *S*, under approximately 23-hr. food deprivation, was placed by hand in the starting box, facing the door. Approximately 1 sec. later the starting-box door was raised. As soon as *S* left the starting box, the door was closed to prevent retracing. Another door located about 4 in. in front of the goal box was lowered after *S* had crossed that point, again to prevent retracing.

A 45-mg. dry pellet was placed, on reinforced trials, in a small porcelain cup in the corner of the goal box. On nonreinforced acquisition trials and on extinction trials neither cup nor pellet was present. Because of a shielding black curtain, *S* was never able to see whether the cup was present until he was well inside the goal box. Approximately 10 sec. after *S* had stopped the clock, and after he had eaten the pellet if so scheduled, he was removed from the apparatus and returned to his living cage to await the next trial. After the scheduled number of trials in the runway, *S* was given his daily food ration, minus the amount he had eaten during experimentation.

On Day 1, *S*s were given 10 pretraining trials with a 1-min. intertrial interval and reinforcement was administered on each trial. On Day 2, 30 acquisition trials were given with either a 15-sec. or a 2-min. intertrial interval and with either 50% or 100% reinforcement, depending

TABLE 1

EXPERIMENTAL DESIGN

Variables	1-min. Spontaneous Recovery		5-min. Spontaneous Recovery		10-min. Spontaneous Recovery		60-min. Spontaneous Recovery		n*	
	Extinction		Extinction		Extinction		Extinction			
	15-sec.	2-min.	15-sec.	2-min.	15-sec.	2-min.	15-sec.	2-min.		
15-sec. acquisition 50% reinforcement	3	2	3	2	3	2	3	2	20	40
2-min. acquisition 50% reinforcement	3	2	3	2	3	2	3	2	20	
15-sec. acquisition 100% reinforcement	3	2	3	2	3	2	3	2	20	40
2-min. acquisition 100% reinforcement	3	2	3	2	3	2	3	2	20	
n*	12	8	12	8	12	8	12	8		
	20		20		20		20			

* Cell values refer to cell n's. Marginal values refer to n's for various rows and columns.

on the group to which *S* was assigned. On Day 3, 30 extinction trials were given, also with either a 15-sec. or 2-min. interval, followed by a spontaneous recovery trial after 1, 5, 10, or 60 min. of delay. The experiment, then, followed a $4 \times 2 \times 2 \times 2$ factorial design with unequal but proportional cell *n*'s. Table 1 shows the design. The marginal labels indicate the experimental variables and the values within the cells indicate cell *n*s.

It took approximately 20 min. elapsed time to complete 10 pretraining trials, about 35 min. for the 30 15-sec. acquisition trials, about 1½-hr. for the 30 2-min. acquisition trials, and 35 min. and 1½ hr., respectively, for the 15-sec. and the 2-min. extinction trials. The spontaneous recovery tests took 1, 5, 10, or 60 min. The possible differences in drive as a result of elapsed experimental times was controlled to some extent by starting *S*s so that the deprivation period for Trial 15, either acquisition or extinction, was constant. There remained, however, a differential deprivation time that was directly related to the spontaneous recovery interval. The maximum difference between *S*s run under massed extinction and 1-min. spontaneous recovery delay and *S*s run under spaced extinction and 60-min. spontaneous recovery delay was about 2 hr. Since all *S*s were run under approximately 23-hr. hunger drive and, in addition, had been on a restricted diet for at least 7 days, it is believed that the differential deprivation period during spontaneous recovery did not affect their behavior markedly. The nonsignificance of the interaction between extinction intertrial intervals and spontaneous recovery partially justifies this belief.

Two percentages of reinforcement (50% and 100%) were used during acquisition, and the 50% reinforcement was presented in two different patterns according to criteria stated by Sheffield (**18**). The two patterns were randomly assigned to the appropriate conditions and subsequent analysis showed no significant difference between them.

Performance in both acquisition and extinction was recorded in units of .01 sec. Then medians of groups of three trials were computed for each *S* and the means of these medians transformed to common logarithms. Analyses of variance and covariance were then performed on the log scores. Each of the five points graphed in the figures is the mean of the logs of two of the median running times, and thus represents a block of six trials.

TABLE 2

SUMMARY OF ANALYSIS OF VARIANCE FOR ACQUISITION DATA

Source	*df*	Mean Square	*F*
Reinforcement	1	.333	6.06*
Acquisition	1	.010	
Reinforcement × acquisition	1	.037	
Error	76	.055	

* Significant at the 5% level.

RESULTS

Acquisition.—An analysis of the logs of the medians for the last five pretraining trials showed no significant differences among the groups which were later to be treated differently. As Table 2 indicates, the only significant difference in acquisition is due to the percentage of reinforcement. Those animals that were rewarded 100% of the time ran significantly faster than those that were rewarded only 50% of the time. The differences between the two groups are significant for all 30 trials and for the last 15 trials. The intertrial interval had no apparent effect on running time, nor was there a significant interaction between reinforcement percentages and intertrial intervals.

Extinction.—The log of the mean of the last two three-trial medians for acquisition was used as a control for an analysis of covariance of the extinction data. The results of this analysis are shown in Table 3 where it can be seen that running times were significantly affected by percentage of reinforcement. The group which had received 50% reinforcement ran faster than the group which had received 100% reinforcement.

The acquisition intertrial interval also had a significant effect upon the extinction running times. Those animals running under massed (15-sec.) acquisition ran faster during extinc-

TABLE 3

SUMMARY OF ANALYSES OF EXTINCTION DATA

Source of Variation	Analysis of Variance			Analysis of Covariance		
	df	Mean Square	*F*	*df*	Mean Square	*F*
Reinforcement	1	.1834	3.24	1	.3639	7.35**
Acquisition interval	1	.3551	6.27*	1	.2552	5.16*
Extinction interval	1	.0910	1.61	1	.1611	3.26
Reinforcement × acquisition	1	.0234		1	.0157	
Reinforcement × extinction	1	.1397	2.47	1	.2362	4.77*
Acquisition × extinction	1	.0544		1	.0481	
Reinforcement × acquisition × extinction	1	.0287		1	.0174	
Error	72	.0566		71	.0495	

* Significant at the 5% level.
** Significant at the 1% level.

tion than did *Ss* who learned under spaced (2-min.) conditions.

The analysis, based on all 30 extinction trials, shows no significant difference as a function of massed vs. spaced extinction trials. If, however, only the last 15 trials are used, the *F* is 4.50 which exceeds the 3.98 which is required for significance for 1 and 71 *df* at the .05 level. The *Ss*, then, who were extinguished with spaced trials ran significantly faster during later trials than those extinguished with massed trials. This is shown quite clearly in Fig. 1.

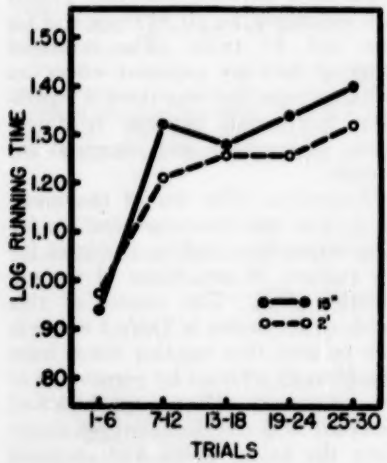

FIG. 1. Extinction curves showing the relationship between log running time and successive blocks of extinction trials with extinction intertrial interval as a parameter. Each point represents the mean of the logs of two three-trial medians. The 15-sec. curve is based on 48 *Ss*. The 2-min. curve is based on 32 *Ss*.

The nonsignificant *F* for the interaction between percentage of reinforcement and acquisition interval can be interpreted to mean that partial reinforcement has its effect independently of the acquisition interval. That this may not be the whole story seems to be indicated by Fig. 2, in which the P-15[2] group appears to extinguish slower than the other groups. This observation is not confirmed by the *t* ratio of 1.92 between the P-15 and P-2 groups. The *t* between the P-15 and all other groups combined, however, was 2.28, which is significant at the .03 level for 78 *df*.

There was a significant interaction between percentage of reinforcement and the extinction interval. This interaction is interpreted to mean

[2] P-15 refers to the condition of partial reinforcement with a 15-sec. acquisition interval. C-2 refers to continuous reinforcement with a 2-min. acquisition interval; 15-15 refers to 15-sec. acquisition interval and 15-sec. extinction interval.

TABLE 4

SUMMARY OF ANALYSES OF SPONTANEOUS RECOVERY DATA

Source of Variation	Analysis of Variance			Analysis of Covariance		
	df	Mean Square	*F*	*df*	Mean Square	*F*
Reinforcement	1	.0405		1	.0244	
Acquisition interval	1	.4004	2.14	1	.0963	
Extinction interval	1	.0275		1	.0735	
Spontaneous recovery time	3	.5588	2.99*	3	.6575	3.83*
Reinforcement × acquisition	1	.0069		1	.0205	
Reinforcement × extinction	1	.5536	2.97	1	.6108	3.56
Reinforcement × spontaneous recovery time	3	.1888		3	.1723	1.00
Acquisition × extinction	1	.4738	2.54	1	.5931	3.46
Acquisition × spontaneous recovery time	3	.8749	4.69**	3	.9174	5.34**
Extinction × spontaneous recovery time	3	.2458	1.317	3	.1530	
Reinforcement × acquisition × extinction	1	.0304		1	.0512	
Reinforcement × acquisition × spontaneous recovery time	3	.2355	1.26	3	.2952	1.720
Reinforcement × extinction × spontaneous recovery time	3	.0537		3	.1376	
Acquisition × extinction × spontaneous recovery time	3	.0904		3	.0640	
Reinforcement × acquisition × extinction × spontaneous recovery time	3	.0231		3	.0210	
Error	48	.1867		47	.1717	

* Significant at the 5% level.
** Significant at the 1% level.

that the resistance to extinction following 50% reinforcement is greater when the extinction interval is 2 min. than when it is 15 sec.

There is no significant interaction between acquisition and extinction intervals, indicating that no over-all statistically significant amount of generalization decrement occurred. Figure 3, however, shows that after spaced acquisition, massed extinction intervals resulted in faster extinction than spaced intervals. There is very little difference between massed and spaced extinction intervals after massed acquisition. The triple interaction between percentage of reinforcement, acquisition interval, and extinction interval also was not significant.

The number of trials for each animal at or below the over-all median extinction running time (15.08 sec.) was determined in a manner similar

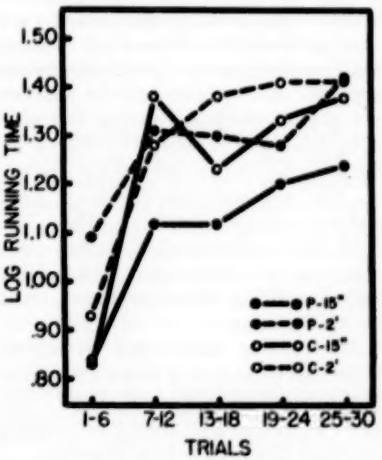

FIG. 2. Extinction curves showing the relationship between log running time and successive blocks of extinction trials with percentage of reinforcement and acquisition intervals as parameters. Each point represents the mean of the logs of two three-trial medians. Each curve is based on 20 Ss.

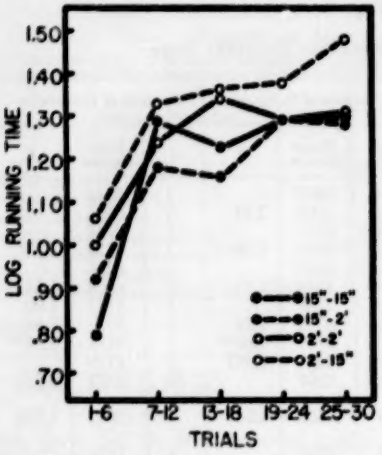

FIG. 3. Extinction curves showing the relationship between log running time and successive blocks of extinction trials with acquisition and extinction intervals as parameters. Each point represents the mean of the logs of two three-trial medians. The 15-15 and 2-15 curves are based on 24 *Ss* each, and the 15-2 and 2-2 curves are based on 16 *Ss* each.

to that reported by Sheffield (**18**) and Stanley (**22**), and an analysis of variance performed on these data. There were no changes in significance levels between this analysis and the analysis of log-mean-median running times.

Spontaneous recovery.—Table 4 indicates that the analysis of covariance of the spontaneous recovery data yielded only two significant *F*'s. The *F* of 3.83 indicates that spontaneous recovery is a function of the delay period, and Fig. 4 indicates the nature of this function.

The other significant *F* shows spontaneous recovery to be a function of the acquisition interval, the 15-sec. interval resulting in more spontaneous recovery than the 2-min. interval.

DISCUSSION

Extinction

Partial reinforcement and the acquisition interval.—Jenkins and Stanley (**10**) cite a considerable number of experiments which present evidence, with few exceptions, that a partially reinforced response is more resistant to extinction than one which is continuously reinforced. Sheffield, following Hull, reasoned that this superiority was a function, in part, of the intertrial interval. With short intertrial intervals, nonreinforced trials are frequently followed closely by reinforced trials, which have the effect of conditioning the response to cues normally present only during extinction. If the intertrial interval is great, nonreinforced trials will not be followed closely by reinforcement and no superiority in resistance to extinction should result. Sheffield showed experimentally that this analysis held for her situation; a response partially reinforced combined with a 15-sec. intertrial interval was more resistant to extinction than a continuously reinforced one with a 15-sec. interval, but when the interval was increased to 15 min. the superiority of the partially reinforced response disappeared.

The very significant *F* of Table 3

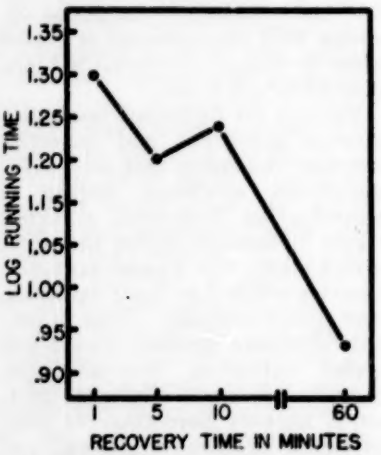

FIG. 4. Spontaneous recovery curve showing the relationship between log running time on the first postextinction trial and the time interval between the last extinction trial and the first postextinction trial. Each point represents the log running time of 20 *Ss*.

again provided support for the notion that a partially reinforced response is superior in resistance to extinction to a continuously reinforced one, but the nonsignificant interaction between reinforcement and acquisition interval is contrary to Sheffield's notion. It is apparently in agreement with a similar nonsignificant interaction found by Grant, *et al.* (6) in an eyelid conditioning study, and with a recent report by Weinstock (25) showing partial reinforcement to increase resistance to extinction even when a 24-hr. interval separates acquisition trials. Figure 2, however, indicates that the group that received partial reinforcement with massed intervals extinguished slower than the other three groups. The significance of this difference was only partially confirmed by *t* tests.

Table 2 also indicates that the acquisition interval significantly affects resistance to extinction, and Fig. 3 shows that it is the massed acquisition interval which leads to superior resistance to extinction. This same result was found by Grant *et al.*, and perhaps it may be explained by appealing to Appelzweig's (2) notion that a more effortful response results in better learning.

Massed vs. spaced extinction trials and frustration drive.—The *F* for the effects of the extinction interval for all 30 trials is not significant, but it is when only the last 15 trials are analyzed. Figure 1 shows that this difference between the massed and the spaced extinction groups indicates quicker extinction with massed trials. Theories of extinction making principal use of an inhibition factor have been considered to require that massed extinction trials result in quicker extinction than spaced. Much evidence indicates this to be the case. Sheffield, however, reports 15-min. intertrial interval *S*s to extinguish faster than 15-sec. *S*s. Stanley (22), using a running-time measure, found confirmation for this position and adapted a frustration-drive theory of extinction to account for it. The present study, with an experimental situation patterned after Sheffield's, found evidence directly contrary to hers but in line with previous findings. The only important procedural difference between the present study and that of Sheffield appears to be the length of the spaced intertrial interval. Her interval was 15 min.; the present interval was 2 min. It may well be that the function relating distribution of extinction trials and extinction running times is not a monotonic one.

A frustration-drive notion (1, 17, 22) typically maintains that ordinary extinction increases the frustration-drive state of the organism and that, for massed extinction trials, the frustration-produced drive does not have time to dissipate between trials. Massed-trial *S*s, therefore, run under greater drive than do *S*s with spaced trials.

There are a number of theoretical and empirical objections to such a notion. If frustration drive is assumed to operate in the same manner as *D* within Hull's theory, which is a common assumption, then the effect of increasing *D* is to increase the strength of all responses in the repertoire at the moment. If the "correct" response should be high in the hierarchy, then increasing *D* will increase it proportionately more than the others. If there are no considerable differences among responses, then increasing *D* will increase competition among responses and decrease the probability of any given response (23). At the beginning of extinction we can assume that the running response is high in the hierarchy of responses and therefore the frustration drive resulting from massed intervals will increase its probability of occurrence disproportionately to that of the others. Later in the extinction trials, however, the strength of the running response will decrease, by whatever assumed mechanism, eliminating any large difference between it and competing responses. This means that extinction curves should show a considerable superiority for massed trials early in extinction and that this superiority should diminish and perhaps disappear later in extinction. There is no evidence for these effects in the curves of Sheffield and Stanley.

Skinner (20) has shown that the ex-

tinction curves after periodic reconditioning are much smoother than those after continuous reinforcement, and he has characterized the postcontinuous reinforcement cyclic fluctuations as emotional. If this is to say that extinction following continuous reinforcement produces more frustration drive than extinction following partial reinforcement, then one would have to predict, other things being equal, that massed trials following continuous reinforcement would lead to slower extinction than massed trials following partial reinforcement. There is no evidence for this in Sheffield's data, and data from the present experiment indicate that spaced trials following partial reinforcement result in the most resistance to extinction.

Miller and Stevenson (11) in an experiment on extinction and spontaneous recovery specifically observed "agitated behavior" which they considered to be an index of emotionality. They observed that on successive extinction days both agitated behavior and running times increased. In their experiment, emotionality as indexed by agitated behavior coincided with a slowing up of running speed.

It seems safe to conclude, then, that there is evidence that nonreinforcement does lead to frustration, but the evidence is far from conclusive that this frustration serves to decrease running time.

Spontaneous recovery.—Table 4 and Fig. 4 present evidence on spontaneous recovery. Only the data from one trial subsequent to extinction was used in the analysis of covariance, and the log mean of the medians for the last two groups of three extinction trials were used for equating all groups at the end of extinction.

Spontaneous recovery did occur and as a function of time. The other significant F of Table 3 shows that postextinction scores were a function of the acquisition interval. This finding can be explained on the basis that a more effortful response is better learned (2) and that spontaneous recovery is a function of acquisition habit strength (26).

Summary

White rats were run in a 4-foot alley runway with running time taken as the response measure. Four variables were manipulated factorially: 15- sec. and 2-min. intertrial intervals in acquisition, 50% and 100% reinforcement, 15-sec. and 2-min. intervals in extinction, and 1-, 5-, 10-, and 60-min. delay intervals before a spontaneous recovery test. The results showed:

1. During acquisition, running times were significantly faster for 100% reinforcement than for 50% reinforcement.

2. Following 50% reinforcement, running times in extinction were significantly faster than following 100% reinforcement.

3. The 15-sec. "massed" acquisition interval led to faster running times in extinction than did the 2-min. "spaced" interval.

4. The massed extinction interval led to significantly slower running times for the last 15 extinction trials than did the spaced interval, although this difference did not attain statistical significance for all 30 trials.

5. Partial reinforcement led to significantly greater resistance to extinction if followed by a spaced extinction interval than if followed by massed.

6. The interaction between percentage of reinforcement and the acquisition interval was not significant.

7. Spontaneous recovery was found to be a significant function of the postextinction delay interval.

8. Spontaneous recovery was significantly greater following a massed acquisition interval than a spaced.

9. The relevance of some of these findings to important theoretical notions was indicated.

References

1. Amsel, A., & Roussel, J. Motivational properties of frustration. I. Effect on a running response of the addition of frustration to the motivational complex. *J. exp. Psychol.*, 1952, **43**, 363–368.
2. Applezweig, M. H. Response potential as a function of effort. *J. comp. physiol. Psychol.*, 1951, **44**, 225–235.
3. Calvin, J. S. Decremental factors in conditioned-response learning. Unpublished doctor's thesis, Yale Univ., 1939.
4. Ellson, D. G. Quantitative studies of the interaction of simple habits. I. Recovery from specific and generalized effects of extinction. *J. exp. Psychol.*, 1938, **23**, 339–358.
5. Ellson, D. G. Spontaneous recovery of the GSR as a function of the recovery

interval. *J. exp. Psychol.*, **25,** 1939, 586–600.

6. Grant, D. A., Schipper, L. M., & Ross, B. M. Effect of intertrial interval during acquisition on extinction of the conditioned eyelid response following partial reinforcement. *J. exp. Psychol.*, 1952, **44,** 203–210.
7. Hilgard, E., & Marquis, D. G. Acquisition, extinction, and retention of conditioned lid responses to light in dogs. *J. comp. Psychol.*, 1935, **19,** 29–58.
8. Humphreys, L. G. Distributed practice in the development of the conditioned eyelid reaction. *J. gen. Psychol.*, 1940, **22,** 379–389.
9. Humphreys, L. G. The effect of random alternation of reinforcement on the acquisition and extinction of conditioned eyelid reactions. *J. exp. Psychol.*, 1939, **25,** 141–158.
10. Jenkins, W. O., & Stanley, J. C., Jr. Partial reinforcement: a review and critique. *Psychol. Bull.*, 1950, **47,** 194–234.
11. Miller, N. E., & Stevenson, S. S. Agitated behavior of rats. *J. comp. Psychol.*, 1936, **21,** 205–231.
12. Pavlov, I. P. *Conditioned reflexes* (Trans. by G. V. Anrep) London: Oxford Univ. Press, 1927.
13. Porter, J. M., Jr. Extinction of an acquired response as a function of the interval between successive non-reinforced elicitations. *J. comp. Psychol.*, 1938, **26,** 261–270.
14. Porter, J. M., Jr. Experimental extinction as a function of the interval between successive non-reinforced elicitations. *J. gen. Psychol.*, 1939, **20,** 109–134.
15. Reyna, L. J. Experimental extinction as a function of the interval between extinction intervals. Unpublished doctor's thesis, State Univ. of Iowa, 1946.
16. Reynolds, B. Extinction of a trace conditioned responses as a function of the interval between successive nonreinforced elicitations. *J. gen. Psychol.*, 1945, **35,** 91–95.
17. Rohrer, J. H. Experimental extinction as a function of the distribution of extinction trials and response strength. *J. exp. Psychol.*, 1947, **37,** 473–493.
18. Sheffield, V. F. Extinction as a function of partial reinforcement and distribution of practice. *J. exp. Psychol.*, 1949, **39,** 511–526.
19. Sheffield, V. F. Resistance to extinction as a function of distribution of extinction trials. *J. exp. Psychol.*, 1950, **40,** 303–313.
20. Skinner, B. F. *The behavior of organisms: an experimental analysis.* New York: D. Appleton-Century, 1938.
21. Spence, K. W., & Norris, E. B. Eyelid conditioning as a function of the intertrial interval. *J. exp. Psychol.*, 1950, **40,** 716–720.
22. Stanley, W. C. Extinction as a function of the spacing of extinction trials. *J. exp. Psychol.*, 1952, **43,** 249–260.
23. Taylor, J. A., & Spence, K. W. The relationship of anxiety level to performance in serial learning. *J. exp. Psychol.*, 1952, **44,** 61–64.
24. Teichner, W. H. Experimental extinction as a function of the intertrial intervals during conditioning and extinction. *J. exp. Psychol.*, 1952, **44,** 170–178.
25. Weinstock, S. Resistance to extinction of a running response following partial reinforcement under widely spaced trials. *J. comp. physiol. Psychol.*, 1954, **48,** 318–322.
26. Youtz, R. E. P. Reinforcement, extinction, and spontaneous recovery in a non-Pavlovian reaction. *J. exp. Psychol.*, 1938, **22,** 305–318.

(Received January 24, 1955)

Journal of Experimental Psychology
Vol. 51, No. 1, 1956

INTERRELATIONSHIPS OF SUCCESSIVE AND SIMULTANEOUS DISCRIMINATION [1]

ALVIN J. NORTH
Southern Methodist University
AND MALCOLM JEEVES
Cambridge University

Spence (**5**) has proposed a three-level theory of discrimination learning, namely, that response tendencies differentiate to elements, compounds, and transverse patterns of cues. In addition, Spence holds that a higher-order cue will not acquire excitatory tendencies toward approach in a problem in which lower-order ones are differentially reinforced—a "minimum level" hypothesis. He says that compounding occurs "only when no one of the cue members is systematically reinforced more than the others" (**5**, p. 90). Likewise, he asserts that transverse patterning takes place "only under conditions that would not permit learning on the basis of a single reinforced component or some simpler type of cue-position pattern (e.g., the type of patterning involved in the successive problem)" (**5**, p. 92).

The major purpose of the present study was to test the above "minimum level" hypothesis, which has been challenged by Bitterman (**2**). Table 1 shows the habits that would be acquired in standard simultaneous and successive white-black discrimination problems if learning occurred on all three levels. Spence holds that the simultaneous problem is solved on the basis of color elements, and the successive one by means of color-position compounds. If so, then if the successive discrimination were learned after simultaneous discrimination, response tendencies to colors would become equated; and if simultaneous were learned after successive, response tendencies to two compounds would become *reversed*. In either case, retention of the first-learned problem and subsequent concurrent performance of both problems would become impaired or impossible.

Another purpose of the experiment was to test Lawrence's hypothesis (**4**) of acquired distinctiveness of cues by examining transfer from simultaneous to successive discrimination, and conversely, varying the degree of learning of the first-learned problem.

METHOD

Subjects.—The *S*s were 32 male albino rats bought from the Charles River Laboratories. They were about 90 days old when the study began.

Apparatus.—The apparatus has been described in detail elsewhere (**1**). It was a four-unit maze with a pair of painted gates side by side in each unit. In the simultaneous task, the white gate was on the right in two units and on the left in the other two, in a varied order. In the successive task, a pair of black gates appeared in two units and a pair of white ones in the other two units, again in a varied order. Color and position preferences were counterbalanced. The negative gate in each unit was locked.

Pretraining.—The rats were placed on a 20- to 22-hr. hunger drive schedule, which continued throughout the study. Pretraining consisted of a series of steps ending with a limited number of forced choices to counteract position preferences in which gray gates were used. Wet mash was used for reward.

Design.—The 32 rats were divided at random into two equal groups. Group I learned simul-

[1] This study was part of a series comprising the "Cognition Project" under the direction of Professor Jerome S. Bruner.

TABLE 1

EXAMPLE SHOWING HYPOTHETICAL HABITS AT EACH LEVEL IN THE LEARNING OF SIMULTANEOUS AND SUCCESSIVE WHITE–BLACK DISCRIMINATIONS

Level of Cues	Simultaneous Condition		Successive Condition	
	Stimulus Unit	Response Tendency	Stimulus Unit	Response Tendency
First level: elements	W	approach W	W	nondifferential
	B	avoid B	B	
	R	nondifferential	R	nondifferential
	L		L	
Second level: compounds	WL	approach WL	WL	approach WL
	WR	approach WR	WR	avoid WR
	BL	avoid BL	BL	avoid BL
	BR	avoid BR	BR	approach BR
Third level: transverse patterns	WL-BR	approach WL (or W)		
	WL-BR	avoid BR (or B)		
	BL-WR	avoid BL (or B)		
	BL-WR	approach WR (or W)		
			WL-WR	approach WL
			WL-WR	avoid WR
			BL-BR	avoid BL
			BL-BR	approach BR

taneous first and successive second, while Group II learned successive first and simultaneous second. Each main group was further divided into two subgroups of 8 *S*s. Subgroups IA and IIA learned their first problem to the criterion of two successive sets of 20 correct out of 24 choices. Subgroups IB and IIB *overlearned* their first task by 120 and 192 trials beyond the criterion, respectively. More overlearning trials were given on the successive task since a pilot study had shown it to be more difficult.

After each rat had learned its first task, it was trained on the second one, which it learned to the criterion or for 192 trials (48 runs), which-

TABLE 2

ERRORS UNDER SUCCESSIVE AND SIMULTANEOUS CONDITIONS

Group	Simultaneous (1st Task)		Successive (2nd Task)			
	Errors to Criterion		Errors to Criterion		Errors in 192 Trials	
	Mean	*SD*	Mean	*SD*	Mean	*SD*
IA	10.75	6.30	33.12	21.02	40.25	16.67
IB	8.88	4.14	21.88	14.50	28.89	10.99
I	9.81	5.41	27.50	18.91	34.56	15.22
	Successive (1st Task)		Simultaneous (2nd Task)			
IIA	56.12	17.66	17.75	12.59	21.38	11.13
IIB	63.50	21.36	13.88	8.87	14.62	10.40
II	59.81	19.94	15.81	11.06	18.00	10.70

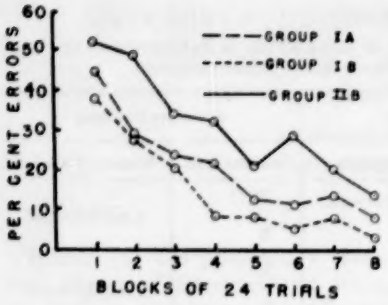

FIG. 1. Learning curves for successive discrimination.

ever was the more. After that, all rats were retrained for 24 trials (6 runs) on the second-learned task and then for 24 trials (6 runs) on the first-learned task.

Finally, all rats were tested in one 32-trial (8-run) session for concurrent mastery of both types of discrimination. In each run, two stimulus arrangements were of the simultaneous type and two of the successive type, in a restricted random order.

RESULTS

The successive problem was more difficult than the simultaneous one (see Table 2). Every rat in Group II made more errors than did the slowest learner in Group I. Since the two groups differed significantly in variance ($P < .01$), the nonparametric Mann-Whitney test was used. The difference in sum of ranks was significant ($P < .01$).

A second result was positive transfer from simultaneous to successive discrimination (see Table 2 and Fig. 1). Comparing errors to criterion in learning the successive task, Group IA saved 45% and Group IB 54% as compared with Group II. Comparing errors in the first 192 trials by analysis of variance, Subgroups IIB and IA differed significantly ($P < .05$); so did IIB and IB ($P < .01$), but IA and IB did not.

A third result was some evidence of *negative* transfer from successive to simultaneous discrimination (see Table 2 and Fig. 2). In the following tests of significance, error scores were subjected to the square-root transformation. Comparing errors in the first 192 trials of the simultaneous task by means of analysis of variance, Subgroup IIA differed significantly from IB ($P < .01$); other subgroup differences, however, were nonsignificant. According to Fig. 2, Subgroups IIA and IIB did not reach quite as high a level of performance as did IB. An analysis of variance of (transformed) error scores summed over Trials 25–192 inclusive yielded the following levels of significance: IB vs. IIA, $P < .01$; IB vs. IIB, $P < .01$); and IIA vs. IIB, $P > .05$.

When transferred to the successive task, Group I *Ss* initially responded more slowly and with more vacillation (as indicated by anticipatory approaches and withdrawals) to previously negative than to previously positive gates. This effect was greatest in the initial trials (being on the order of several minutes difference in latency for some rats), disappearing by about the third 24-trial session.

In retraining the rats in their second-learned task, 97.4% of their responses were correct. Retention of the first-learned discrimination was

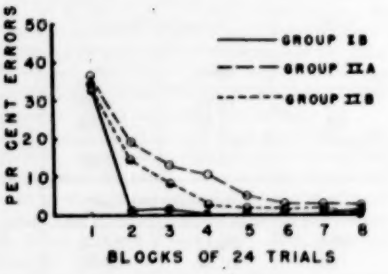

FIG. 2. Learning curves for simultaneous discrimination.

shown by the fact that 85.8% of the responses were correct in retraining.

In the final test for concurrent mastery of successive and simultaneous problems, 89.7% of the responses were correct. Of the 32 rats, 24 had only four or fewer errors, while 14 had only one or zero. The mean per cent correct was 88.9 for the successive trials, and 90.6 for the simultaneous ones.

Discussion

The results are contrary to the previously discussed implications of Spence's "minimum level" hypothesis to the effect that higher-order cues will not acquire response tendencies in a problem in which lower-order cues are differentially reinforced. Retention of each first-learned discrimination after learning of the second one, and final concurrent mastery of both problems imply *transverse patterning from the outset.* Table 1 shows that habits at this level would account for these results.

As an alternative to Spence's "minimum level" hypothesis the following *multiordinal* one is proposed: All orders of functional stimulus units—elements, compounds, transverse patterns, etc.—*concurrently* acquire response tendencies. *Which* order or orders of stimulus unit are functional in a given problem is an empirical matter, depending probably on "structural" factors in perception, as emphasized by gestalt psychologists, and also on prior learning, such as posited by the hypothesis of acquired distinctiveness of cues. Presumably the rate of differentiation of response tendencies at the various levels would depend on the relative discriminability of the several cues.

This multiordinal hypothesis appears to account for most of our results. It explains Group I's differential latencies in the initial trials of the successive task on the basis of transferred responses to color elements, and the subsequent reduction of these differences as the result of nondifferential reinforcement. Transverse patterning accounts for the results previously mentioned. Moreover, the hypothesis accounts for the fact that the simultaneous problem was learned more readily than the successive one, since in the former differentiation of response tendencies would occur at all *three* levels (see Table 1), particularly at the highly learnable color-element level, although evidence (3) shows that the simultaneous transverse patterns were less distinctive than the successive ones.

The status of compounding, however, is problematic. If compounding habits were influential, then the following effects, none of which occurred, would be predicted from Table 1: (*a*) negative transfer from simultaneous to successive discrimination; (*b*) in the initial trials of the successive task, Group II would have responded more rapidly to previously positive compounds than negative ones; and (*c*) pronounced interference between simultaneous and successive discriminations during the test for concurrent mastery. The only evidence for compounding was Group II's negative transfer, which might have been caused by persisting position habits acquired in the difficult successive task. All in all, if compounding occurred at all, only weak color-position habits developed.

In view of the foregoing results and theoretical analysis, no critical evaluation of the hypothesis of acquired distinctiveness of cues can be made on the basis of this study. Group I's positive transfer is confirmatory, and observed differential responses to colors may be the mediating responses in question. Overlearning gave no significant advantage, although the difference was in the predicted direction. Group II's negative transfer, while not confirming, might have been caused by other factors such as persisting position preferences.

Summary

An experiment was designed to test Spence's three-level theory of discrimination learning and also Lawrence's hypothesis of acquired distinctiveness of cues. Group I (16 rats) learned a simultaneous problem followed by a successive

one, while Group II (16 rats) learned the problems in the converse order. Each group was subdivided so as to provide two degrees of learning of its first problem. After retraining on the second problem and then on the first, all rats were tested for concurrent mastery of both problems.

Successive was more difficult than simultaneous discrimination. There was positive transfer from simultaneous to successive discrimination, but some evidence of negative transfer in the opposite direction. Differences between criterion and overlearning groups were nonsignificant. In the final test, both types of discrimination problem were successfully performed.

Concurrent mastery of both problems was explained as transverse patterning. In contrast with Spence's "minimum level" hypothesis, the following multiordinal one was proposed, and appeared to account for most of the results: All orders of functional stimulus unit—elements, compounds transverse patterns, etc.—*concurrently* acquire response tendencies.

In view of the results and theoretical analysis, the evidence pertaining to Lawrence's hypothesis regarding acquired distinctiveness of cues did not permit critical evaluation of this view.

References

1. Bruner, J. S., Matter, J., & Papanek, M. L. Breadth of learning as a function of drive level and mechanization. *Psychol. Rev.*, 1955, **62**, 1–10.
2. Bitterman, M. E. Spence on the problem of patterning. *Psychol. Rev.*, 1953, **60**, 123–126.
3. Bitterman, M. E., & Wodinsky, J. Simultaneous and successive discrimination. *Psychol. Rev.*, 1953, **60**, 371–376.
4. Lawrence, D. H. Acquired distinctiveness of cues: I. Transfer between discriminations on the basis of familiarity with the stimulus. *J. exp. Psychol.*, 1949, **39**, 770–784.
5. Spence, K. W. The nature of the response in discrimination learning. *Psychol. Rev.*, 1952, **59**, 89–93.

(Received January 31, 1955)

Journal of Experimental Psychology
Vol. 51, No. 1, 1956

ACQUISITION OF A CONDITIONED RESPONSE AS A FUNCTION OF FORWARD TEMPORAL CONTIGUITY

M. E. FITZWATER[1] AND RANDOLPH S. THRUSH[2]

Bowling Green State University

Early investigators, such as Bernstein (**1**) and Wolfle (**9**), concerned with the problem of the relationship between the generation of a CR and the temporal order and disparity between the CS and the UCS, used as indices of learning either the total number of CR's elicited in all test trials given throughout training or the number of CR's elicited in test trials given at the completion of training. Their conclusions were that various forward, simultaneous, and backward time relationships would produce conditioning differing only in the amount or strength of conditioning.

More recently, however, the contiguity problem has been attacked with different emphasis. Consistent with the study of a learning problem is the study of learning curves. Therefore, the problem has been reopened with emphasis upon the shapes of the acquisition curves under various temporal conditions. Kimble (**4**) found that .1- and .2-sec. interonset intervals of forward conditioning produced extremely low incidence of conditioning, approximately equally distributed in occurrence throughout training; while .3- and .4-sec. intervals produced the conventional negatively accelerated learning curves. He used the classical eyeblink situation.

Spooner and Kellogg (**7**), using the instrumental finger withdrawal to shock, demonstrated that simultaneous and backward interonset temporal relationships did not yield the usual learning function. Responses to the CS were predominant during the early test trials, diminishing in frequency during later test trials. Fitzwater and Reisman (**2**) found that the simultaneous and backward curves were not significantly different from their pseudo-conditioning curve. They substantiated the findings of Spooner and Kellogg in that these curves showed an initial rise during early training but the response tendency dropped to near zero incidence toward the end of training.

From these studies, it seemed highly desirable to determine more precisely the minimal forward temporal interval which is capable of yielding the conventional, negatively accelerated, conditioning curve and if possible to determine at what temporal interval, longer than optimal, the conditioning tendency begins to recede to the pseudo-like function. This then constituted the problem herein investigated.

Method

Subjects.—Thirty male and 30 female beginning psychology student volunteers were used. They were assigned by closed random order to the temporal conditions of .0 or simultaneous, .1, .2, .3, .4, and .6 sec. of forward temporal interonset interval.

Apparatus.—In all groups the instrumental finger-withdrawal-to-shock situation was used. A 1024-cycle tone about 40 db. above threshold was presented to *S*, through headphones, as the CS.

The UCS, a 60-cycle AC shock, was adjusted to the minimal intensity necessary for the consistent withdrawal of the finger from a microswitch lever electrode.

Stimuli were controlled by means of an electronic timing apparatus readily adjustable to the different modes of presentation. A chronometer

[1] Now at The RAND Corporation, Santa Monica, California.

[2] Now at The Ohio State University.

was started with the onset of the first stimulus and was stopped when *S* withdrew his finger. The *S* was in an adjoining room to that of *E* with two-way microphone contact between them. To minimize voluntary behavior *S*s were provided with reading material and were instructed that the experiment was designed to test the effects of different types of distraction upon reading rate and comprehension, and that they would be given a reading test later. Subsequent questioning after the administration of the reading test indicated that none of the *S*s were aware of the conditioning aspect of the study.

Procedure.—A pretest of three tonal stimulations was given to each *S* to determine the original neutrality of the tone. If *S* made a finger withdrawal to the tone, he was excused from the experiment. The tones were followed by several presentations of the shock to determine the minimal, consistent response level and to accustom *S* to this stimulus.

The *S* was then given a total of 100 trials—80 reinforced training trials and 20 test trials (one randomly placed in each five trials).

In all groups the tone was presented for .2 sec. and the shock was presented for .2 sec. or until the finger response occurred, if it was less than .2 sec. All trials were presented at one sitting. Responses were counted as CR's if they occurred after the onset of the CS but before an elapsed time of .7 sec.

Results

In order to ascertain whether or not an initial inequality existed between the groups, a chi square was run on the number of CR's obtained in each group on the first test trial. This chi square was found to be 8.89 (11.07 required for $P = .05$, $df = 5$). It was therefore concluded that the groups differed initially within the limits of chance.

For the purpose of further analysis and for curve plotting the test trials were grouped into Trials 1–4, 5–8, 9–12, 13–16, and 17–20. Figure 1 shows the frequency of CR's in each of the temporal groups during each test trial interval.

Data for Trials 1–8, 9–20, and 1–20 were treated by analysis of variance. The *F* ratio was found to be significant for all three sets of combined trials, with Trials 1–8 significant at the 5% level ($F = 2.38$), Trials 9–20 significant at the 1% level ($F = 5.99$), and Trials 1–20 also significant at the 1% level ($F = 5.77$). Each *F* was interpreted with 5 *df* between groups and 54 *df* within groups.

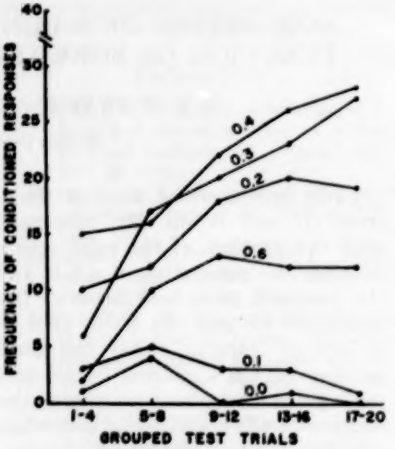

Fig. 1. Conditioning curves showing the frequency of CR's in successive groups of four test trials for 10 *S*s in each interonset interval group.

Since *F* for Trials 9–20 was significant, individual *t*'s were computed between each possible pair of groups for these trials. The following six *t*'s were significant at the 1% level of confidence: between the .0-sec. and the .2-, .3-, and .4-sec. groups; and between the .1-sec. and the .2-, .3-, and .4-sec. groups. No other *t* approached significance at even the 5% level. The .6-sec. group was neither significantly greater than the .0- or .1-sec. groups, nor was it significantly less than the .2-, .3-, or .4-sec. groups.

Discussion

The results with the simultaneous and the .4-sec. groups were in accord with the findings of previous studies, both yielding performance similar to that occurring in comparable groups of Spooner and Kel-

logg (7) and of Fitzwater and Reisman (2).

It appears that an interonset interval in excess of .1 sec. is necessary in order to establish conditioning. Moreover, though further work appears necessary, beginning with a time interval of .6 sec., the amount of conditioning appears to be diminished. The study of longer time intervals should be undertaken to establish the minimum time at which the curve again will approximate that of the pseudo or simultaneous function.

These results generally are in agreement with those of Moeller (6) which indicated a substantial decrement in conditioning for 1.0- and 2.5-sec. intervals using the Féré GSR.

Contiguity as it applies to the interonset interval between the CS and the UCS may be a relatively "narrow band" phenomenon and therefore it seems of utmost importance to work within the range of this band, which appears safely to extend from .2 sec. through .5 sec., whenever other variables are being investigated.

The optimal interonset interval has been found to be approximately the same regardless of whether the reaction latency is short, as in the present study, or long, as in studies by White and Schlosberg (8) and Moeller (6) which used the GSR. It appears that stimulus asynchronism of from .2 to .5 sec. is necessary for conditioning and that neural asynchronism may be inferred as necessary at central association levels. Studies such as those by Harlow and Bromer (3) and Light and Gantt (5) have already demonstrated that conditioning is a central process not requiring the elicitation of the motor reactions which become elicitable.

Summary

Six groups of 10 *Ss* received 80 training and 20 interspersed test trials under the following CS-UCS interonset intervals: .0-, .1-, .2-, .3-, .4- and .6-sec. Results indicated that with the .0- and .1-sec. intervals the finger-withdrawal response was not conditioned to the tone. The conventional acquisition curve was obtained under intervals of .2, .3, and .4 sec. The .6-sec. group results showed less acquisition than these groups but more than the .0- and .1-sec. groups (not significantly different from either of the above functions).

References

1. Bernstein, A. L. Temporal factors in the formulation of conditioned eyelid reactions in human subjects. *J. gen. Psychol.*, 1934, **10**, 173–197.
2. Fitzwater, M. E., & Reisman, M. N. Comparison of forward, simultaneous, backward, and pseudo-conditioning. *J. exp. Psychol.*, 1952, **44**, 211–214.
3. Harlow, H. F., & Bromer, J. A. Acquisition of new responses during inactivation of the motor, premotor, and somesthetic cortex in the monkey. *J. gen. Psychol.*, 1942, **26**, 299–313.
4. Kimble, G. A. Conditioning as a function of the time between conditioned and unconditioned stimuli. *J. exp. Psychol.*, 1947, **37**, 1–15.
5. Light, J. S., & Gantt, W. H. Essential part of reflex arc for establishment of conditioned reflex. Formation of conditioned reflex after exclusion of motor peripheral end. *J. comp. Psychol.*, 1936, **21**, 19–36.
6. Moeller, G. The CS-UCS interval in GSR conditioning. *J. exp. Psychol.*, 1954, **48**, 162–166.
7. Spooner, A., & Kellogg, W. N. The backward conditioning curve. *Amer. J. Psychol.*, 1947, **60**, 321–344.
8. White, C. T., & Schlosberg, H. Degree of conditioning of the GSR as a function of the period of delay. *J. exp. Psychol.*, 1952, **43**, 357–362.
9. Wolfle, H. M. Conditioning as a function of the interval between the conditioned and the original stimulus. *J. gen. Psychol.*, 1932, **6**, 80–101.

(Received January 13, 1955)

Journal of Experimental Psychology
Vol. 51, No. 1, 1956

EFFECTS OF MAGNITUDE OF TEMPORAL BARRIERS, TYPE OF GOAL, AND PERCEPTION OF SELF

CHARLES M. SOLLEY AND ROSS STAGNER

University of Illinois

Discussions of clinical psychology and the psychology of personality abound in assertions that behavior in experimental situations is in part a function of personality. A substantial number of these statements derive from ex post facto views of data, rather than from design. We report here an investigation in which an important personality variable, the individual's perception of himself as "valuable," is incorporated into an experimental design.

For logical reasons (omitted here because of space limitations) we have selected barrier behavior as a suitable context for this purpose. It is commonly asserted that the organism mobilizes additional energy when faced with a barrier, and that behavior at barriers often involves decreased efficiency in skilled tasks. Each proposition falls in an area which might readily be affected by personality variables.

Barriers may be either spatial or temporal. Miller (6) has pointed to the need for additional research on nonspatial (temporal) barrier situations. One way of studying the effects of temporal barriers would be to present *S* with an anagram which has no solution, so that *S* is prevented, with complete blockage for a fixed period of time, from reaching his goal (the solution word). By presenting anagrams which can be solved, then impossible anagrams (temporal barriers), and then more anagrams approximately equal in difficulty to the pre-barrier anagrams, one can study the effects of barriers. In the present study the authors were interested in the effects of magnitude of barriers, type of affectivity associated with material being used, and perception of "value" of self on three measures: (*a*) a cathected motivation measure (palmar sweating), (*b*) spontaneous comments made in the presence of the barriers, and (*c*) solution time for anagrams subsequent to the barriers.

The following hypotheses were proposed for testing. (The logical rationale is presented by Solley [13] in detail.) It was first hypothesized that *S*s faced with several insoluble anagrams would show an increment in palmar sweating (energy mobilization at a barrier), whereas *S*s encountering zero insoluble anagrams would show a decrement. Our second hypothesis was that there would be an increase in solution time for anagrams subsequent to the temporal barriers. Related to this hypothesis was our prediction that amount of increase in subsequent solution times would be partially dependent upon how "valuable" *S* perceived himself to be, i.e., there would be a significant interaction between perception of value of self and magnitude of barriers with respect to these solution times. Also with respect to perception of "value" of self was our fourth hypothesis that *S*s who perceived themselves to be "valuable" would emit spontaneous comments with reference to the task at hand, while *S*s who perceived themselves to be "nonvaluable" would emit comments with reference to "self."

The experimental design was prepared to incorporate another experimentally independent factor. Solution words for anagrams can be classified meaningfully as to whether they are "negative-

affective" (e.g., "enemy"), "positive-affective" (e.g., "beauty"), or "neutral" (e.g., "chair"). It was hypothesized that in the pre-barrier phase of the experiment, where *S* would have the least amount of knowledge of what class of words he was dealing with (3), the order of solution time, from highest to lowest, would be "negative-affective," "neutral," and "positive-affective," respectively. This constituted our fifth hypothesis.

Similarly, it was thought that "negative-affective" materials would be more susceptible to disruption than neutral and "positive-affective" materials. Consequently, our sixth hypothesis was that there would be a significant interaction between magnitude of temporal barriers (number of insoluble anagrams) and "type" of goal (solution word).

It should be pointed out that we did not attempt to state all possible hypotheses, e.g., how palmar sweating operates as a joint function of perception of value of self and magnitude of barriers, since we felt there was too little empirical data to serve as a basis for plausible prediction.

Method

Design

There were two "phases" in the experiment. Phase 1 was the before-insoluble-anagram phase and Phase 2 was the after-insoluble-anagram phase. Three levels of temporal barriers were used between Phase 1 and Phase 2—either 0 (control), 3, or 5 insoluble anagrams in succession with a 2-min. time limit for each anagram, soluble or insoluble. A given *S* received the same type of solution word in Phase 2 as in Phase 1. The three types of solution words were "positive-affective," "neutral," and "negative-affective." For a given phase we had, in effect, a 3-by-3 factorial design.

Three dependent variables were analyzed. Palmar sweating, as measured by a fingerprint stain technique (2, 4, 8) was used to study changes in "motivation." "Spontaneous" comments were recorded to see if they reflected how the *S* perceived himself. Finally, solution time for the seventh anagram—the anagram immediately following the temporal barriers—was analyzed to determine the effects of type of solution word, magnitude of temporal barriers, and perception of "value" of self on barrier behavior.

Subjects

Ten *S*s were assigned to each independent cell of the design for a total of 90 *S*s. Subjects were randomly assigned to conditions as they arrived at the experimental room. All *S*s were volunteers and were undergraduates at the University of Illinois. No attempt was made to control distribution of males and females among the cells of the design, though a χ^2 check showed no significant difference in distribution.

Materials

Materials used in the experiment consisted of (*a*) mimeographed copies of 20 of Osgood's Semantic Differential Scales, 12 of which had high factor loadings on a "valuation" factor (9), (*b*) materials for a fingerprint test of palmar sweating, (*c*) anagrams typed on 3 × 5 index cards, (*d*) record sheets for recording solution times and *S*'s ratings of degree of affectivity of solution words, and (*e*) a standard electric timer. Since some of these materials are not widely known, we describe their basic features in more detail.

Osgood's Semantic Differential Scales.—These scales consisted of two antipodal terms with the numbers 1 through 7 equally spaced between. For example,

Good 1 2 3 4 5 6 7 Bad

In the study by Osgood (9), 50 such scales were factored, yielding three factors—"valuation," "activity," and "potency." In the present study we were interested only in the scales heavily loaded on the "valuation" factor.

Each *S* was given a copy of this test with instructions to check the number on each scale that he thought best described himself as an individual. He was also told that "4" was "average." As soon as *S* had finished checking a number on each of the 20 scales, *E* scanned the protocol to make sure that each scale had one check. An *S*'s score was obtained by summing the numbers checked on the 12 "valuation" scales.

As to the reliability of this measure, it may be noted that Osgood (9) reports that the reliability for ratings of abstract concepts was .85. The present writers found a test (administration to class from which volunteers were obtained)-retest (individual experimental-situation testing) reliability of .87, the second test being given approximately two months later. In the analyses to be reported here, the second test measure was used for categorizing *S*s.

Fingerprint Stain Test of Palmar Sweating.—The basic history and development of this test have been adequately covered by Light (4). Briefly, the measure has received validation

TABLE 1

SOLUTION WORDS AND ANAGRAMS IN THE ORDER PRESENTED FOR EACH TYPE OF SOLUTION WORD IN THE ANAGRAM TASK

Position in Order	Type of Solution Word					
	Positive-Affective		Neutral		Negative-Affective	
	Word	Anagram	Word	Anagram	Word	Anagram
1	escape	cesape	eight	geiht	enemy	neemy
2	beauty	aebuty	double	udoble	cannot	nacnot
3	truth	urtth	corner	orcner	death	aedth
4	sweet	ewset	chair	ahcir	afraid	rfaaid
5	sleep	elsep	center	necter	black	albck
6	dance	nadce	arrive	rraive	yellow	leylow
7						
7	desire	sedire	amount	omaunt	refuse	feruse
8	music	sumic	trade	ratde	blood	olbod
9	enjoy	jneoy	square	uqsare	suffer	fusfer
10	value	lavue	single	nisgle	sudden	dusden
11	accept	cacept	pound	oupnd	spirit	ipsrit
12	peace	aepce	count	ucont	storm	tosrm

through comparison of the measure of patients in therapy ("anxiety neurotics") with the measures of "normal" people (**4**). It has been shown to *decrease* with therapy (**4**) and under relaxing conditions such as listening to soft music (**2**). It has also been shown to increase in "stress" situations (**2**).

A thorough description of the chemical nature of the test may be found in Mowrer (**8**). The test consisted of swabbing some anhydrous ferrous chloride solution on the right or left index finger (depending on whether *S* was right- or left-handed, respectively). After the solution dried for 30 sec., *S* placed this finger on a small piece of paper, treated with tannic acid solution, on top of a small postal scale, holding pressure constant at 8 oz. for 2 min. It has been shown (**11**) that, as the finger sweats, more ferrous chloride is deposited on the paper, causing it to darken. One then "reads," on a densitometer (**8**), the amount of light impedance by change in microampere output of a photoelectric circuit containing a microammeter.

The Anagram Test.—Holding frequency of occurrence of solution words constant at the second 500 most frequent printed English words (**14**), *E* had 10 judges pick out all "affective" words and indicate whether each was "negative" (avoidance) or "positive" (approach) in nature. If words were not picked as "affective," they were considered "neutral."

Restrictions were then placed on these judged populations of words. Only five- or six-letter words were to be used. Secondly, only words which could not be made into other words by rearranging the letters were used. (In practice, it was impossible to check this completely. The weaker meaning of having no alternative solution was used, i.e., probability near 1.00 that *S* would find no alternative solution.)

Having made these restrictions, we drew random samples of 12 words from each of the three "type" groups. This method of sampling insured more generality for any conclusions that might be drawn concerning these populations of words. Having drawn our samples of words, we constructed anagrams by rearranging the first two or three letters of each word, pilot studies having shown that "degree of disorder" of letters was a major variable to be controlled. This procedure for building anagrams insured that they would be solved quite rapidly, i.e., usually less than 10 sec., under nonbarrier conditions. The 36 solution words and their corresponding anagrams are shown in Table 1.

The insoluble anagrams were *putah*, *noogme*, *rookan*, *abnul*, and *scharke*. Although several *S*s remarked that these reminded them of words, they never found an acceptable solution.

Instructions

Each *S* was given three anagrams built from the words *horse*, *bread*, and *clock*. He was allowed to solve these anagrams as practice. The *S* was told that he would be allowed 2 min. to solve each anagram and that after that time the next anagram would be presented. He was to keep working for the entire 2 min. if necessary. The *S* was also told that *E* could not give him the answer to any anagram if he failed to solve it (until the end of the experiment). The *E* then

answered all questions pertaining directly to what *S* was to do.

Procedure

The *S* was brought into the experimental room and allowed to relax a few minutes while *E* took down such things as name, college, and major field of study. Then *S* filled out the Semantic Differential Scales. Following this, *E* gave the fingerprint stain test, telling *S* that it was a measure of how much energy he had available for "thinking." The *S* was then given the instructions for the anagram test, with the practice anagrams. Six soluble anagrams followed, these being followed by either 0, 3, or 5 insoluble anagrams depending on which barrier level *S* was assigned. Following the insoluble anagrams, six more soluble anagrams were presented one at a time. After the 12th (last) anagram, *E* took the fingerprint measure again.

The *E* recorded solution time for each anagram to the nearest .01 sec. although these time measures were accurate only to the nearest .1 sec. All spontaneous comments were recorded as nearly verbatim as possible.

Results

Fingerprint Measures of Palmar Sweating

Frequency plots of the distributions of measures of initial palmar sweating showed no marked deviation from a normal distribution (though there was some skewness) and Bartlett's test for homogeneity of variance yielded a nonsignificant χ^2. There were no significant *F* values for any of the analysis of variance components.

A "change" score was then computed by subtracting initial from final fingerprint score. A constant of 20 was added to all scores to make them positive. For this analysis *S*s in each cell of the design were categorized as to whether they were above or below the grand median for perception of value of self. In this way a high (H) and low (L) perceived-value-of-self dimension was added to the design. A χ^2 showed no significant deviation from equality of number of *S*s from cell to cell on this dimension.[1] Similar tests were made for all interaction cells with the same result. Again, assumptions of normality and homogeneity were checked and found tenable with respect to "change" scores for palmar sweating. The results of the analysis of variance are shown in Table 2.

TABLE 2

Analysis of Variance on Changes in Fingerprint Measure of Palmar Sweating

Source	*df*	Mean Square	*F*	*P*
Words (W)	2	.11	—	
Barriers (B)	2	74.45	4.59	.05
Self Percept (S)	1	21.86	1.35	
W × B	4	25.94	1.60	
W × S	2	63.73	3.93	.05
B × S	2	24.65	1.52	
W × B × S	4	60.20	3.71	.01
Within	72	16.23		
Total	89			

Type of solution word showed no significant relation to change in palmar sweating, nor was the interaction between type of goal word and magnitude of barrier significant. However, the *F* between levels of barriers was 4.59, significant at the .05 level. Further, the average changes in palmar sweating were − 1.28, .73, and 1.95 (microamperes change in output of photoelectric circuit) for 0, 3, and 5 insoluble anagrams, respectively. This is the order predicted by our first hypothesis. The control level did show a *decrement* in palmar sweating whereas the 3 and 5 insoluble anagram levels showed *increments*. By *t* test, only the zero condition differs significantly from the other two.

[1] Since the number of cases per cell was not equal, the analysis was done following the "Method of Weighted Squares of Means" outlined by Snedecor (12, pp. 293–301). It will be noted that Snedecor recommends this test when the interaction term is significant, as was true in the present case.

Although it was not stated as an hypothesis for this experiment, one might reasonably expect that *S*'s perception of his value might be related to changes in palmar sweating. This was not the case. Table 2 shows that the *F* value between high and low perception of value-of-self groups was not significant. One might also expect, as an initial hunch, that there would be a significant interaction between how people perceived themselves and magnitude of barrier effect. Again, this expectation was not confirmed by the analysis of variance shown in Table 2.

Perhaps the most intriguing aspects of changes in palmar sweating are the significances of the interaction between self percept and type of word ($P = .05$) and the interaction between self percept, type of word, and magnitude of barrier ($P = .01$).

Figure 1 summarizes these interactions in graphical form. In the middle portion of Fig. 1 it can be seen that "negative-affective" materials produce opposite results for *S*s with high and low self-value percepts. The *S*s who perceive themselves as low in value show marked increases in palmar sweating with increases in number of insoluble anagrams, whereas *S*s who perceive themselves as high in value show a slight decrement. For "neutral" and "positive-affective" materials, both classes of *S*s tend to react alike, although *S*s with low self-value percepts tend to show less increase in palmar sweating with increased magnitudes of barriers. These results will be more fully discussed later.

Solution-Time Scores

Pre-barrier phase.—Since the effects of individual words were confounded by having a constant order of presentation for each *S*, the solution-time sums for the first six anagrams were used as scores for this phase. A log transformation sufficiently normalized

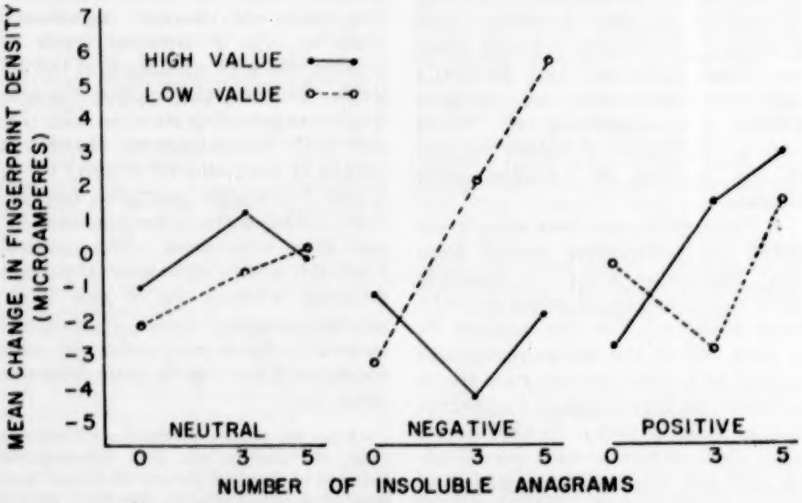

FIG. 1. Relation between evaluation of self and amount of palmar sweating produced by 0, 3, and 5 insoluble anagrams, for neutral, positive, and negative solution words.

TABLE 3

ANALYSIS OF VARIANCE OF LOG TOTAL SOLUTION TIME FOR SIX ANAGRAMS BEFORE BARRIERS

Source	*df*	Mean Square	*F*
Word type (W)	2	.6085	6.56**
Barriers (B)	2	.0354	0.38
W × B	4	.0965	1.04
Within	81	.0928	
Total	89		

** $P = .01$.

the distributions of these scores and Bartlett's test for homogeneity yielded a nonsignificant χ^2. The results of the analysis of variance for this phase are shown in Table 3.

The *F*, with respect to barriers, was nonsignificant, as one would expect since at this point this variable had not been operative. Similarly, the interaction between barrier level and type of solution word was nonsignificant. However, as predicted in Hypothesis 5, there was a significant difference between types of solution word ($P = .01$). The order of the mean log solution times from longest to shortest was "negative-affective" (1.62), "neutral" (1.36), and "positive-affective" (1.33). Since controls had been taken to equate these types of words as to frequency of occurrence in printed English, length of word, and "degree of disorder" of the letters of the anagram, and since "random" samples were drawn from the judged populations of these words, it seems reasonable to conclude that this significant difference was due to the direction of the affectivity of the word.

Postbarrier phase.—Solution time for Anagram 7, the anagram immediately following either 0, 3, or 5 insoluble anagrams, should be the most sensitive postbarrier measure of barrier behavior. A simple log transformation sufficiently normalized the distributions and Bartlett's test for homogeneity of variance showed no significant difference. The results of the analysis of variance on this measure are shown in Table 4.

Table 4 shows no significant differences between types of solution word, and no significant interaction between type of solution word and magnitude of barriers. This result means that our sixth hypothesis has not been confirmed. The most important measure in this analysis relates to the effects of number of insoluble anagrams, which is significant at the .001 level. The mean log solution times for the seventh anagram were .5545, 1.5307, and 1.5136 for 0, 3, and 5 insoluble anagrams, respectively. For 0 as compared with either 3 or 5, the *t* test is significant at the .01 level. The difference between 3 and 5 is not significant.

Analyses with Respect to Perception of Value of Self

Since perception of value of self can be considered an index of perceived homeostatic efficiency in adaptation to stress situation, this variable was also included in the analysis of solution time scores for Anagram 7. The *Ss* were categorized as to whether

TABLE 4

ANALYSIS OF VARIANCE OF LOG SOLUTION TIME FOR SEVENTH ANAGRAM

Source	*df*	Mean Square	*F*
Barriers (B)	2	5.2093	17.59***
Word type (W)	2	.7399	2.50
Self percept (S)	1	.0050	.001
B × W	4	.2354	.79
B × S	2	5.6052	18.92***
W × S	2	.1323	.45
B × W × S	4	.1065	.36
Within	72	.2962	
Total	89		

*** $P = .001$.

they were above or below the grand median "self-value" score and, again, χ^2 tests of deviation from equality of frequencies in different cells of the design were made and found nonsignificant.

The results of the analysis of variance with respect to this variable and its interaction with other dimensions of the basic design are also shown in Table 4. Although there was no overall significant difference in log solution times for *Ss* with high and low self-value percepts, the interaction between high-low perception of value of self and magnitude of barriers was significant at the .001 level of significance. Thus, we conclude that Hypothesis 3 was confirmed.

Hypothesis 4, that *Ss* with high self-value percepts would emit spontaneous comments with reference to the task, whereas *Ss* with low self-value percepts would emit comments with reference to self, was next tested. Thirty-five *Ss* made comments in the presence of the insoluble anagrams. Three judges independently classified these comments, all three judges agreeing on the classification of 34 of the 35 protocols. These 34 were then analyzed. Two-by-two contingency tables—high-low self-value percepts by word-self comment orientation—were set up for each cell of the original design. Fisher's (1) exact method was used to calculate a pooled exact χ^2. This procedure yielded a χ^2 of 22.72, 12 *df*, significant at the .02 level. As predicted, *Ss* who perceived themselves as highly valuable made more comments such as "Is this a word? Is it English?", whereas *Ss* who perceived themselves to be less valuable made more comments such as "I must be stupid."

Discussion

Our first hypothesis predicted an increase in motivation (energy mobilization as measured by palmar sweating) upon encountering a temporal barrier. Our results substantiated this prediction. To the extent that this result can be generalized to other situations, this means that an increase in motivation of the organism is probably a significant parameter of the barrier situation. It might be noted that Wright (15) found similar results in quite different types of barrier situations—thus lending additional support to such a generalization.

The exact consequences of such an increase in motivation stimulus intensity would be dependent upon other operating factors. For example, in the present study a severely restricted external environment was presented, reducing the possibility of increased sampling of external environmental stimulus sources. This restriction could only lead to an increase in scanning of previously established perceptual groupings (13). In other experiments where such a restriction was not present in the design, one might obtain considerably different results.

Speculations are defensible concerning the empirical result that high and low self-value percepts interacted significantly with type of affectivity of materials with respect to changes in palmar sweating. Here we observed that *Ss* with low-value self percepts showed marked increases in palmar sweating when working, under barrier conditions, with "negative-affective" materials, whereas *Ss* with high-value self percept showed a decrease (Fig. 1). It was also noted that *Ss* with low-value percepts were more susceptible to increases in palmar sweating under barriers. We may speculate as to the reasons for these results as follows.

In the developmental, or learning, process, individuals associate stimulus cues with barrier conditions. An individual may continue reacting primarily to the goal, or he may begin operating in terms of the associated or cathected-motivation stimulus cues. The individual who perceives himself as low in value is probably reacting largely to the barrier or "stress" component of our experiment, whereas the individual who

perceives himself to be valuable is reacting more to the cues relevant to identifying the solution words. Admittedly this is speculation, but it is a type of speculation which is amenable to future experimental testing.

The change in solution time as a function of encountering a temporal barrier was taken to be an indirect index of an increase in scanning of alternative, previously learned perceptual groupings at *S*'s command. The increase in motivational intensity is accompanied by physiological changes (energy mobilization, palmar sweating). These changes feed back proprioceptive stimuli to the central nervous system. Inasmuch as perceptual groupings can be considered as a sampling of the total influx of stimuli at a given time, the efficiency decrement may be treated as an interference effect derived from these afferent impulses. To the extent that these stimuli have been cued to perceptions of words, these words would be "wrong" in terms of the demands of the experiment. By greatly increasing the total number of solution words to be considered, this process reduces the probability of locating the correct solution word. Thus delay in reporting the "correct"word is a plausible outcome. The results of our experiment confirmed this prediction of an increase in solution time as function of the occurrence of temporal barriers. This result involved a "mean" prediction for it was recognized that how the individual perceived himself would determine his relative disruption.

It is hypothesized that perception of value of self is an index of homeostatic efficiency of behavior in that it indicates an individual's past success in compensating efficiently for delay in goal-reaching. Analogically (for the present) we may speak of the individual perceiving himself as a tool, making possible speedy or slow solution to problems. Probably such self perceptions arise, genetically, as a consequence of successful dealings with barriers. Two basic types of mechanism (not necessarily discontinuous) appear to operate when the person encounters a barrier. Either he goes "out of the field" (consequently decreasing efficiency) or he attends more closely to the task at hand. Support of this interpretation was found in the analysis of "spontaneous" comments made while the insoluble anagrams were present. Our *S*s who perceived themselves to be "high" in valuation made task or word-oriented comments, while *S*s who perceived themselves to be less valuable made more self-oriented comments, showing that they were no longer attending closely to the task. For them, cues deriving from somatic processes may have been dominant.

With respect to solution times for anagrams as a function of direction of affectivity, our results showed that "negative-affective" (avoidance) type words took the longest to solve, "neutral" words were intermediate, and "positive-affective" words were shortest. Although, strictly speaking, solving an anagram is not the same as seeing a word presented tachistoscopically (as in visual threshold experiments), there is a great deal of functional similarity. For *S* to solve the anagram the solution word must somehow be represented at threshold value. The presentation is "memorial" rather than sensory, but a threshold is still logically involved. When the solution word is presented visually at threshold value, then *S* reports it. The same is true of anagrams.

Rather than postulating some mechanism such as "perceptual defense" (5) which seems to entail *S*'s perceiving the word but withholding it, the present study assumes that *S* has learned avoidance sign values for certain classes of words and approach sign values for other classes. It is these approach and avoidance characteristics which determine speed of "seeing" the solution word.

Summary

An experiment involving temporal barriers was designed. Three levels of temporal barriers —0, 3, and 5 insoluble anagrams in succession— were used with three classes of word goals— "negative-affective," "neutral," and "positive-affective" solution words of anagrams—to form a 3×3 factorial design. Whether *S* had a "high" or "low" valuation of self was hy-

pothesized as an important determinant of barrier behavior. The results may be summarized as follows:

1. Motivational intensity, as measured by increases in palmar sweating, increases with increases in barrier magnitude; however, only the differences between occurrence and nonoccurrence of a barrier are statistically reliable.

2. The direction and magnitude of motivation changes are joint functions of perception of value of self, magnitude of barrier, and type of affectivity of materials.

3. The interaction between perceived value of self and magnitude of barriers, with respect to solution time for postbarrier anagrams, is significant, meaning that barrier behavior varies in part according to how valuable the individual perceives himself to be.

4. There is a disruption of average efficiency in solving anagrams with occurrence of a temporal barrier.

5. The *Ss* who perceive themselves as valuable make task- or word-oriented spontaneous comments in the presence of barriers, whereas *Ss* who perceive themselves as less valuable make more self-oriented comments.

6. The sign of affectivity—positive, neutral, or negative—of goal word is an important determinant of efficiency in solving anagrams under nonbarrier conditions.

References

1. Fisher, R. A. *Statistical methods for research workers.* New York: G. E. Stechert, 1946.
2. Gladstone, R. An investigation of the relationship between palmar sweating and emotion as measured by a group test of palmar sweating. Unpublished doctor's dissertation, Univ. of Illinois, 1949.
3. Lacey, O. W., Lewinger, N., & Adamson, J. F. Foreknowledge as a factor affecting perceptual defense and alertness. *J. exp. Psychol.*, 1953, **45**, 169–174.
4. Light, B. Tension changes in patients undergoing psychotherapy. Unpublished doctor's dissertation, Univ. of Illinois, 1951.
5. McGinnies, E. Emotionality and perceptual defense. *Psychol. Rev.*, 1949, **56**, 244–251.
6. Miller, N. E. Experimental studies of conflict. In J. McV. Hunt (Ed.), *Personality and the behavior disorders*, Vol. 1. New York: Ronald, 1944.
7. Mood, A. M. *Introduction to the theory of statistics.* New York: McGraw-Hill, 1950.
8. Mowrer, O. H. *Psychotherapy—theory and research.* New York: Ronald, 1953.
9. Osgood, C. F. The nature and measurement of meaning. *Psychol. Bull.*, 1952, **49**, 197–237.
10. Selye, H. Experimental evidence supporting the concept of "adaptation energy." *Amer. J. Physiol.*, 1938, **123**, 758–765.
11. Silverman, J. J., & Powell, V. E. Studies on palmar sweating. *Psychosom. Med.*, 1944, **6**, 243–249.
12. Snedecor, G. W. *Statistical methods.* Ames, Iowa: Collegiate Press, 1946.
13. Solley, C. M. Drive, barrier conditions, and personality variables affecting problem-solving behavior. Unpublished doctor's dissertation, Univ. of Illinois, 1954.
14. Thorndike, E. L., & Lorge, I. *The teacher's wordbook of 30,000 words.* New York: Teachers College, Columbia Univ., 1944.
15. Wright, H. F. The influence of barriers upon strength of motivation. *Contrib. Psychol. Theor.*, 1937, **1**, No. 3. Durham, N. C.: Duke Univ. Press.

(Received January 3, 1955)

Journal of Experimental Psychology
Vol. 51, No. 1, 1956

THE ESTIMATION OF LOUDNESS BY UNPRACTICED OBSERVERS[1]

S. S. STEVENS AND E. C. POULTON[2]

Psycho-Acoustic Laboratory, Harvard University

The relation between subjective loudness and the physical intensity of the stimulating tone has been investigated by a wide variety of methods yielding a wide variety of results (**3, 6**). The decibel change in the stimulus required to produce a 2:1 ratio in apparent loudness has been shown to cover a range of the order of 20 db. Many factors seem to influence the judgment of *O* when he attempts to estimate relative loudnesses in a quantitative manner, and the instability of these judgments has led some workers to despair of being able to specify a loudness scale that will prove valid for a typical listener. Others prefer to believe that the difficulties encountered in loudness judgments serve only to enhance the interest of the problem, and that by discovering the sources of the perturbations in loudness estimations it may prove possible to eliminate or control them. The work reported here was undertaken in this latter spirit.

The judgment of loudness ratios is at best a delicate, difficult business and many of the experimental techniques that have been applied to the problem have been loaded with sources of potential error. Biases seem to creep in whenever an unwary experimenter subjects a listener to subtle suggestion by word, by procedure, or even by the nature of the apparatus. It is our belief that many of these biases can be identified and eliminated and that, within the limits of normal human variability, a representative loudness scale can be constructed.

We present evidence that 65 unpracticed *O*s, on their *first* judgments, were able to produce consistent determinations of loudness ratios under two different procedures: a method of adjustment and a method of magnitude estimation. These methods are two of several we have tried to explore in the course of a larger program in which we have sought to detect and isolate some of the factors that influence loudness judgments. They are good methods in that they are potentially capable of leaving *O* free of all constraints other than those imposed by the stimuli themselves, but in practice both methods have potential sources of distortion that need to be discovered and controlled.

An unconstrained *O* is, of course, an absolute necessity. Garner (**1**) demonstrated nicely how constraints can twist the results of loudness experiments when he showed that under the method of constant stimuli the value called "half as loud" can be shifted as much as 20 db by controlling the range of the variable stimuli presented to *O*. The effects of most constraints and biasing factors are not so dramatic as this, but they may nevertheless be important.

[1] This research was carried out under Contract N5ori-76 between Harvard University and the Office of Naval Research, U. S. Navy (Project NR142-201, Report PNR-163). Reproduction for any purpose of the U. S. Government is permitted. We are indebted to W. S. Verplanck, E. G. Heinemann, and M. J. Moskowitz who kindly arranged for their students to serve as *O*s.

[2] Now returned to the Applied Psychology Research Unit, Cambridge, England.

Garner proved that loudness experiments can be designed in such a way that context is highly controlling. The same has been shown by Stevens, Rogers, and Herrnstein (**8**). In our present experiments, by limiting our concern to first trials, we try to minimize context and thereby permit *O* to make a judgment that is relatively free of this particular influence.

Observers.—A total of 76 *Os* (36 men and 40 women) were used in these experiments. Most of them were students at Harvard or Radcliffe. None of them had previously made loudness judgments and all were ignorant of the purpose of the experiments. For the experiments using the method of adjustment, 44 *Os* were divided into four groups of 11 *Os*. For the magnitude estimations, 32 *Os* were divided into four groups of eight *Os*.

Experiment I: Attenuator Comparison

In Exp. I we try to evaluate (*a*) bias introduced into the method of adjustment by the characteristics of the device used by *O*, and (*b*) effect of certain procedural details on adjustments to loudness. The data cited are drawn from a larger experiment of rather complicated design in which an attempt was made to sample several variables. Only the relevant parts are presented here.

Method

Apparatus and procedure.—The output of a 1000-cycle oscillator was led to two push-type switches. The left-hand switch presented the standard loudness (about 90 db re .0002 dyne/cm²) and the right-hand switch presented the variable loudness. After passing through a narrow 1000-cycle filter (General Radio 830-R) the tone was presented to *O* binaurally through a pair of PDR-8 earphones in sponge-Neoprene cushions (MX-41/AR). The *O* varied the loudness by means of either a decibel attenuator or a "sone potentiometer." This sone potentiometer consisted of two tapered units (General Radio 371-T) connected as shown in Fig. 1. The decibel attenuator provided 120 1-db steps in its 360° rotation. It had no detents. The relation between the position of the knobs of these two controls and the attenuations produced is also shown in Fig. 1.

The purpose of the sone potentiometer is simply to provide *O* with a control whose position is more nearly proportional to loudness than is the position of the control on a conventional decibel attenuator. Unlike the decibel attenuator, equal increments of rotation of the sone potentiometer sound approximately like equal increments in loudness. Actually this particular potentiometer departs from the ideal control in the same direction as the decibel attenuator, but it is certainly a better approximation of the ideal.

With the variable controls set for zero attenuation, the maximum level of the variable tone was 4 db higher than that of the standard.

Each *O* made two adjustments to half loudness per week for two weeks. Group I used the decibel attenuator the first week and the sone potentiometer the second week, while Group II used the sone potentiometer the first week and the decibel attenuator the second week. Five *Os* in Group I and six in Group II started with the attenuator set at 94 db both weeks, and then made a second adjustment with the attenuator preset near threshold. The remaining *Os* started with the reverse settings. When comparing the standard and variable, *O* was told to listen first to the standard for about 1 sec., then to pause for about 1 sec., and finally to listen to the variable for about 1 sec.

The treatment of the two groups differed in two other respects. First, the absolute threshold of Group I but not of Group II was measured before adjustments to half loudness on both weeks. When the threshold was measured, the attenuator was preset at 94 db, so that *O* had to turn the attenuator knob all the way down to threshold. The other difference was that on both weeks Group II had to approach the final setting by bracketing, whereas Group I was not permitted to bracket. The *E* checked that *Os* in Group II listened alternately to louder and softer tones before settling on the point of half loudness. The *Os* in Group I had to start again from the extreme attenuator position whenever they overshot the point of ½ loudness and tried to return.

These two different treatments (thresholds and bracketing) were held constant during the comparison of the two types of attenuator. They were afterwards changed in some exploratory sessions devoted to other problems.

Problems of averaging.—Group means were calculated in decibels in the traditional way. But since *O* was actually judging loudness and not decibels, the means have also been calculated by first converting each decibel reading into units of loudness and reconverting the means of these units into decibels. We call this average the "sone average." Its rationale (**5**) is prin-

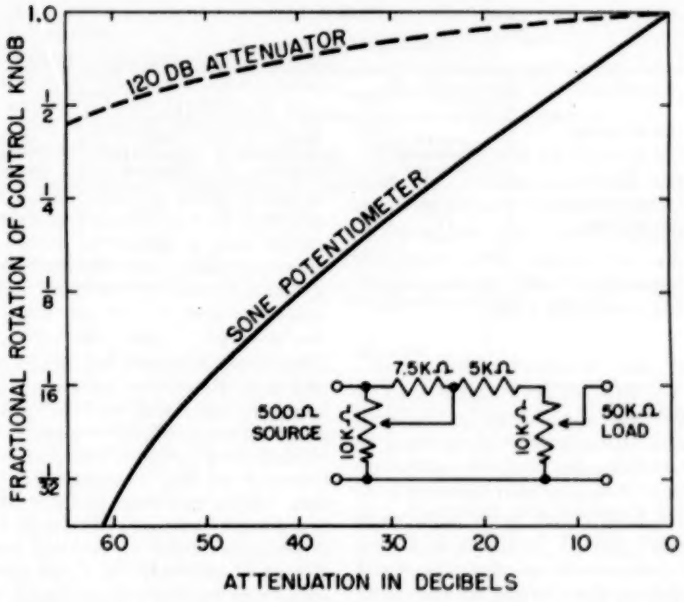

FIG. 1. Diagram and attenuation characteristic of the "sone potentiometer." With this potentiometer in the circuit, the loudness *O* hears is roughly proportional to the position of the control knob. The attenuation characteristic of the decibel attenuator is also shown.

cipally that it undoes some of the skewness that is virtually inevitable in the decibel measures. That this is so is shown by the fact that the sone averages tend to be closer to the medians than are the decibel averages. Improper averaging has been a source of bias in the treatment of the results of many loudness experiments.

The transformation used to get the sone averages was $L = kI^{.33}$, where L is loudness, k is a constant and I is acoustic intensity (power). In choosing this transformation we were being deliberately conservative. The formula represents a loudness scale on which a 12-db reduction corresponds to half loudness.

Separate tests of significance were performed on both sets of averages. The significance levels quoted are always the more conservative. It is a nice question which, if either, of these two sets of averages is suitable for tests of significance.

Results and Discussion

The results are shown in Table 1. On both weeks Group I required more attenuation for ½ loudness than did Group II. This we believe is attributable principally to the fact that for Group I a threshold measurement was made at the beginning of each session. We note, however, that although the two groups were treated rather differently, they both required less attenuation with the sone potentiometer than with the decibel attenuator. Combining the sone averages of the two groups for the two weeks, we find that the decibel attenuator gave 13.3 db attenuation for ½ loudness, and the sone potentiometer gave 10.6 db. This difference of 2.7 db is significant at the .001 level.

Bias due to attenuator.—The 2.7-db difference in the attenuation (sone average) required for ½ loudness, which was produced simply by changing attenuators between Weeks 1 and 2, suggests that these inexperienced *O*s were influenced by the character-

TABLE 1

ATTENUATOR COMPARISON*

Type of Average	Group I		Group II	
	Week 1 (Decibel attenuator)	Week 2 (Sone potentiometer)	Week 1 (Sone potentiometer)	Week 2 (Decibel attenuator)
Decibel average	15.4	12.7**	9.3	12.3***
Sone average	14.5	12.2**	8.9	12.0***
Median	14.0	12.5	8.4	12.5

* Entries show the decibel reductions required to produce ½ loudness. Each group contained 11 *O*s.
** Week 1—Week 2, $p = .05$ with a one-tailed test.
*** Week 1—Week 2, $p = .001$.

istic of the attenuator they used. When the relation between the turn of the knob and the change in loudness produced is extremely nonlinear, *O* seems unable completely to discount the fact. Being at best uncertain of his initial judgments, he perhaps tends to assume that ½ loudness is somewhere near halfway down toward threshold on the control he uses. If this tendency exists, the adjustments with the decibel attenuator would tend to exaggerate the attenuation required for half loudness. This is what appears to have happened. On the other hand, the mean attenuation required for half loudness with the decibel attenuator on the first week was only 14.5 db (Table 1, Group I), which is nowhere near the 40 db that would have been required if knob position were the sole controlling factor. Loudness as it sounded to *O* was clearly the principal determinant of the adjustments—even with the decibel control.

In devising an ideal potentiometer control for an experiment of this sort, we might in principle resort to a kind of successive approximation. Beginning with any arbitrary attenuator, we might find the loudness function it yields, and proceed to use this information to design an improved control, which we would then test and improve in successive stages. We have approximated this procedure in the present experiment. Thus on the first week the *O*s of Group I, who used the decibel attenuator, gave a mean attenuation of 14.5 db. The sone potentiometer, which attenuates about 14 db at its midpoint of turn, corresponds therefore to this loudness-decibel relation. With the sone potentiometer, a new group of unpracticed *O*s gave a mean attenuation of 8.9 db for half loudness (Group II on Week 1 in Table 1). This result may still have been biased by the characteristic of the potentiometer, and ideally another potentiometer should be designed with a midpoint of turn providing about 9 db attenuation. This new control should then be used on another group of unpracticed *O*s. It seems unlikely, however, that the results would be greatly affected by carrying this process so far. Presumably the effect of improving the potentiometer becomes negligible as the ideal is approached.

The precaution of using a new group of unpracticed *O*s whenever the potentiometer is revised may or may not be necessary. In favor of it is the argument that *O*s who have settled on their criterion of half loudness while using one control may carry this criterion into succeeding experiments. In certain other experiments (**8**) we have occasionally obtained evidence of such carry-over. We have also obtained evidence that the type of attenuator used may exert only a slight influence on the adjustments made by highly experienced *O*s.

Other considerations.—The *O*s of Groups I and II were subsequently employed in some additional experiments

that gave incidental evidence on certain procedural matters.

Effects due to "adaptation level" seem to be relatively minor in loudness judgments, but they certainly must be considered, and if possible they should be either counterbalanced or eliminated. In this and other experiments we find evidence, for example, that when *O* listens for a period to very faint levels, as in the course of setting a tone to his absolute threshold, his subsequent settings to ½ loudness are lower than they would otherwise be. It seems reasonable to conclude that if we want to determine the typical quantitative estimates of relative loudness for a group of *S*s, we should not force them to focus their attention on one particular loudness level immediately before they are presented with a very different level.

In the method of adjustment it also seems highly desirable to insist that *O* approach his final setting by successive approximations (bracketing). For example, the effect of the position of the control when *O* starts his adjustment is greater when bracketing is eliminated than when it is insisted upon.

Even with bracketing, however, the starting position on the control presents a certain problem. Although we tried to solve this problem by starting with the variable set alternately at each of two extremes (4 db higher than the standard and near threshold), we do not advocate this procedure. We believe that a better rule is for the experimenter to set the starting position of the variable stimulus in a random manner at various distances above and below the average position at which the *O*s make their final settings. This procedure appears to have worked well in an experiment on the halving and doubling of the loudness of white noise (**2**).

It should also be pointed out that, in using the sone potentiometer, *O* should not be required to make settings near the low end of its range where a very small turn of the knob is sufficient to halve the apparent loudness. The control becomes too coarse in this region. A good working rule is to set the levels so that *O* can use the upper three-quarters of the potentiometer range.

Experiment II: Fractional Loudnesses

This study was performed to determine the extent to which the first loudness adjustments of unpracticed *O*s using the sone potentiometer are consistent with the assumption that on the average the *O*s use a common scale of subjective loudness.

Method

The apparatus and procedure were the same as for Group II on the first week of the previous experiment, except in two respects. First, the 11 *O*s in Group III made their two adjustments to ¾ loudness, the 11 *O*s in Group IV made their two adjustments to ¼ loudness. And secondly, these adjustments were followed at once by a single adjustment to ½ loudness, and then an adjustment to ¼ loudness for Group III, and to ¾ loudness for Group IV. These additional adjustments started with the sone potentiometer set at random somewhere near the middle of its range of movement.

Results

The results are shown in Table 2, together with the results of the first two adjustments made by Group II

TABLE 2

Fractional Loudness*

Loudness Ratio	Type of Average	Attenuation in Decibels		
		Group II	Group III	Group IV
3/4	Decibel		*4.7*	*6.1*
	Sone		*4.5*	5.8
	Median		*3.8*	5.8
1/2	Decibel	*9.3*	*8.6*	*10.6*
	Sone	*8.9*	8.1	9.2
	Median	*8.4*	7.8	8.8
1/4	Decibel		17.0	*19.0*
	Sone		16.0	*15.3*
	Median		14.5	*15.5*

* Entries show the decibel reductions required to produce the fractional loudnesses shown in the first column. Each group contained 11 *O*s. Italicized entries stand for the first judgments made by each group.

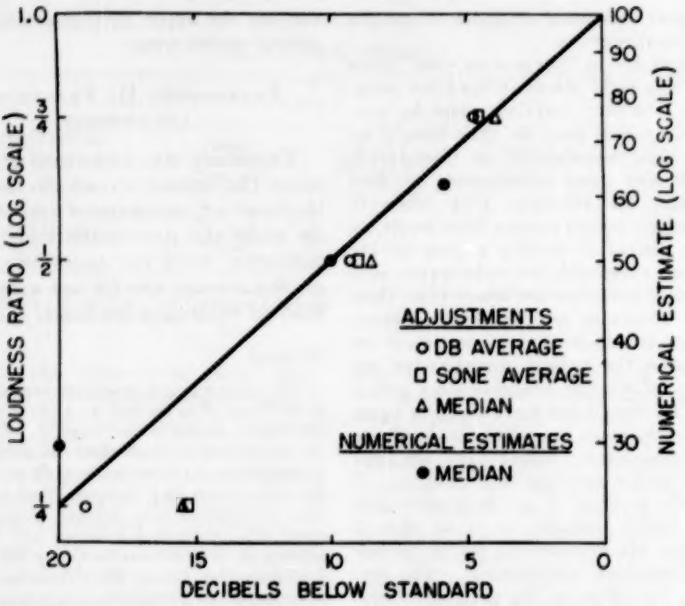

FIG. 2. The unfilled points show the intensities to which *O* set the variable tone to make its loudness stand in various ratios (left-hand scale) to the loudness of the standard. Each point is for a group of 11 *O*s, each of whom made two settings—the first he had ever made. The filled points show the first magnitude estimations (right-hand scale) made by different groups of 8 *O*s each. Each *O* made one estimate. Each point is for a separate group. The straight line with a slope such that 10 db corresponds to ½ loudness represents a portion of the loudness scale derived from previous experiments.

with the sone potentiometer in the previous experiment. The data for the first adjustments of each group have been entered in a log-log plot in Fig. 2. If loudness varies as a power function of intensity, the relation can be represented in this plot by a straight line. The straight line in Fig. 2 is actually a portion of the loudness scale obtained by pooling all the previous results in this field obtained by about a dozen different experimenters (6). Its equation is $L = kI^{0.3}$. The fact that the experimental points approximate a straight line in a log-log plot indicates that the different groups of *O*s were consistently using a common loudness scale—even on their very first judgments.

The data for the first adjustments to the three fractional ratios are as consistent with a power-function loudness scale as are the data obtained when the same *O*s made adjustments to all three ratios (see Table 2). This finding is evidence against the hypothesis that the unpracticed *O* makes his first judgment of loudness more or less at random, and then judges consistently with respect to his first commitment. For if this hypothesis were true, the first adjustments of the different groups of *O*s should have been about the same, whatever the ratio of loudness asked for. The fact that the data line up as well as they do in Fig. 2 suggests that, on the average, unpracticed *O*s in a reasonably unconstrained situation make their first loudness judgments con-

sistent with a power-function loudness scale.

Actually, the slight departures of the adjustment data from the loudness function in Fig. 2 are in a direction consistent with the foregoing hypothesis, in the sense that the $\frac{3}{4}$ adjustment is a bit too low and the $\frac{1}{4}$ adjustment is a bit too high. But the case for a common loudness scale is not necessarily weakened thereby, for it is reasonable to suppose that there is a best ratio to set as the task for *O*. Certainly such extreme ratios as .999 and .001 ought to prove difficult, especially to an *O* making his first adjustment. We might reasonably expect to find that, in adjusting for a ratio that calls for a variable set close to the standard, *O* would tend to overshoot (set too far from the standard), and that in adjusting for ratios far from the standard he would tend to undershoot (set too near to the standard). Slight deviations in these expected directions are what we seem to find in the present experiment.

Which ratio provides the optimal task for *O* is an interesting question. Setting to a ratio far from the standard has the advantage that the results determine the slope of the loudness function over a wide range. On the other hand, when the ratio is such that it requires the variable to be set far from the standard, the task becomes more difficult and systematic biases become more evident. A final answer to this problem is probably not possible on the basis of the available evidence, but from the outcome of other experiments with more-practiced *O*s we are inclined to conclude that consistent determinations of the slope of the loudness function can be obtained from adjustments made at least to such ratios as 1/2, 1/4, 1/10. Evidence from other experiments also indicates that in order to balance out certain systematic biases that occur in the use of fractional ratios, these settings should be complemented by adjustments to multiple ratios, e.g., 2, 4, and 10.

Experiment III: Magnitude Estimation

The consistency of *O*s' first judgments of loudness was also explored by a completely different method (7) in another experiment, the relevant parts of which are here reported.

Method

This method, which we call magnitude estimation, is a very simple procedure: *O* hears a standard tone followed by a variable tone. The *E* assigns a number to the loudness of the standard, e.g., 100, and *O* assigns whatever number he thinks would appropriately describe the variable loudness. The *O* is in no way limited in his choice of numbers. The method has been used by one of us (SSS) for judgments of subjective brightness (4) and subjective weight as well as for loudness.

In the present experiment four groups of eight *O*s each judged a different variable relative to a common standard. The apparatus was similar to that used in the previous experiments. The standard, which was called 100, had a sound-pressure level of about 100 db. The different variables were 6, 10, 20, and 40 db less intense than the standard.

Results

The medians of the numbers assigned to three of these variables on the *O*s' very first trials are plotted in Fig. 2. Space does not permit the plotting of the median estimate given to the variable that was 40 db below the standard. The median estimate for this stimulus was 12. The value we would predict by extrapolating the straight line in Fig. 2 is about 6.

The data plotted in Fig. 2 show that direct estimations of sensory magnitude are reasonably consistent with a power-function loudness scale and that the two methods (magnitude estimation and adjustment) give quite concordant results. Under the conditions of this particular experiment the agreement between the two methods appears to be especially good in the region of $\frac{1}{2}$ loudness.

Discussion

In the previous experiment we noted a slight tendency to adjust the variable a little too far from the standard for the ratio 3/4 and a little too near the standard for the ratio 1/4. This tendency has its analogue in an equally slight tendency to assign to the variable stimuli numerical estimates that are a little large at the low end and a little small at the high end. As is evident in Fig. 2, these two tendencies result in deviations in opposite directions from the average loudness function, and we have here a suggestive example of how the evidence from two procedures might be combined to cancel out systematic errors or biases in each of them.

The fact that in the experiment on magnitude estimation the standard intensity was about 10 db higher than it was in the experiment on adjustment may have contributed to the slightly flatter slope of the function determined by the numerical estimates. In a number of experiments we have found that these slopes become flatter as the level of the standard is increased, but an increase from 90 to 100 db usually makes only a small difference. We might also note in passing that, when *Os*' task is to give numerical estimates to loudnesses *greater* than the standard, the slope gets steeper instead of flatter as the level of the standard is increased (**7**).

Figure 2 suggests that the measurement of loudness is not hopeless—only difficult. If given a proper opportunity, people can make consistent quantitative statements about what they hear. It is true that such factors as context or adaptation level—the influence on one stimulus of others in the series—can be made to have potent effects, as Garner has demonstrated. But it is also true that we can minimize such factors in the experimental design. In our experiment there was no series of stimuli to provide a context, for each *O* heard only one comparison tone. Even under these stringent circumstances, the judgments were clearly consistent with the assumption that loudness is scalable, and that the loudness heard by the typical *O* in a free situation is a power function of intensity.

Summary

These experiments tested the ability of 65 *Os* to make consistent quantitative judgments of the relative loudness of tones on their first trials. They also explored some of the biasing factors that enter such experiments. The method of adjustment and the method of magnitude estimation were used.

Conclusions

1. In the method of adjustment it is advisable to use a "sone potentiometer" on which the turn of the knob is approximately proportional to the loudness produced.

2. On his first trial an unpracticed *O* can adjust the intensity of a tone to produce an arbitrary loudness ratio relative to another tone. Each of three groups of *Os* adjusted the tone to a different loudness ratio in a consistent manner.

3. On their first trial, different groups of *Os* can consistently make direct numerical estimates of the relative loudnesses produced by different intensities.

4. The results for the method of adjustment corroborate those for the method of magnitude estimation, and both approximate the loudness scale derived from previous investigations.

References

1. Garner, W. R. Context effects and the validity of loudness scales. *J. exp. Psychol.*, 1954, **48**, 218–224.
2. Poulton, E. C., & Stevens, S. S. On the halving and doubling of the loudness of white noise. *J. acoust. Soc. Amer.*, 1955, **27**, 329–331.
3. Robinson, D. W. The relation between the sone and phon scales of loudness. *Acustica*, 1953, **3**, 344–358.
4. Stevens, S. S. Biological transducers. *Convention Rec., I.R.E.*, 1954, Part 9, 27–33.
5. Stevens, S. S. On the averaging of data. *Science*, 1955, **121**, 113–116.
6. Stevens, S. S. The measurement of loudness. *J. acoust. Soc. Amer.*, 1955, **27**, 815–829.
7. Stevens, S. S. The direct estimation of sensory magnitudes: loudness. *Amer. J. Psychol.*, in press.
8. Stevens, S. S., Rogers, M. S., & Herrnstein, R. J. The apparent reduction of loudness: a repeat experiment. *J. acoust. Soc. Amer.*, 1955, **27**, 326–328.

(Received January 3, 1955)

Journal of Experimental Psychology
Vol. 51, No. 1, 1956

DISCRIMINABILITY AND STIMULUS GENERALIZATION[1]

NORMAN GUTTMAN AND HARRY I. KALISH[2]

Duke University

Three different forms for the stimulus generalization gradient have been proposed: concavity (Hull, **6**), convexity (Spence, **12**) and linearity (Schlosberg and Solomon, **10**). In the formulation of each of these proposals, however, there has been a tacit or explicit assumption of some relationship between discriminability, in the psychophysical sense, and the generalization decrement. This assumption is manifested in Hull's postulate that generalization gradients are decreasing exponential functions on a j.n.d. scale, and also in Schlosberg and Solomon's proposal that generalization gradients are straight lines on an equal-appearing interval scale. Spence's use of a logarithmic stimulus plot appears to stem from similar considerations.

This assumption does not only appear to be intuitively reasonable, but can also be shown to have a rational basis by comparing the dimensions of the difference threshold and the generalization decrement (from which the generalization gradient can be derived by integration). The difference limen (DL) is ordinarily defined by the slope of the psychophysical function. For example,

$$DL = \frac{\Delta S}{[p(R) = .75] - [p(R) = .50]} = \frac{\Delta S}{\Delta p(R)}$$

On the other hand, the slope of the generalization gradient, i.e., the generalization decrement, is given by $\frac{\Delta p(R)}{\Delta S}$. This implies that the generalization decrement should be proportional to the reciprocal of the DL. It is a way of saying that the organism generalizes to the extent that it cannot discriminate or that generalization is the inverse of discrimination. In essence, this proposal is a reformulation of Schlosberg and Solomon's hypothesis (**10**).

The foregoing analysis may be subjected to empirical verification by investigating a continuum for which the discriminability function is known. For the present experiment, wave length of light was chosen as a dimension along which to test generalization because the DL for wave length ($\Delta\lambda$) is not constant over the spectrum. By an appropriate selection of CS values, therefore, it should be possible to produce a set of gradients whose slopes reflect the characteristics of the $\Delta\lambda$ function. Where $\Delta\lambda$ is small, the gradient should be sharp; where $\Delta\lambda$ is large, the gradient should be relatively flat; and if the CS is fixed at a value where $\Delta\lambda$ is either increasing or decreasing, the generalization gradient should show a corresponding asymmetry. This continuum is also interesting because of the

[1] This study was supported by the Duke University Research Council and by grants M-631C and M-629C from the National Institute of Mental Health, United States Public Health Service. The research program of which this is a part is currently supported by grant MH-1002 from the National Institute of Mental Health, United States Public Health Service. The authors are indebted to Harley M. Hanson, Werner K. Honig, and Mrs. Deborah H. Nickerson for their assistance in this research.

[2] Now at University of Missouri.

qualitative changes which occur within it and which would lead to the expectation, on the basis of subjective experience, that generalization would occur extensively within a hue but would decrease abruptly as the transitions between spectral hues are approached.

The technique of the present study, which utilizes the aperiodically reinforced key-pecking response, permits an examination of the generalization gradients for individual *Ss*. Moreover, this technique makes it possible to investigate the generalization gradients at different levels of response strength during the course of extinction, as well as variations in the generalization gradient attributable to individual differences in response strength. The obtaining of generalization gradients for individual *Ss* in this experiment of an outcome of the fact that aperiodic reinforcement greatly increases resistance to extinction, such that the introduction of a test stimulus for a brief interval during experimental extinction reduces the response strength by a small fraction of its total extent, e.g., in a 30-sec. test, 50 responses may be subtracted from a "reserve" of several thousand.

Method

Subjects.—The *Ss* were 24 experimentally naive pigeons maintained by restricted feeding to 80% of their body weight under ad libitum feeding.

Apparatus.—A modified version of the Skinner automatic key-pecking apparatus was used. The *S*'s compartment consisted of a plywood box 14 in. high, 12 in. wide, and 12 in. long and was situated in an air-conditioned, darkened, soundproof chamber. The roof of *S*'s box was clear Lucite and the floor was a perforated metal grating covered with a sheet of translucent plastic. The upper half of one of the walls was made of aluminum with a ½-in. round aperture at the center 6.5 in. above the floor of the box. The *S*'s key, a rectangle of translucent plastic (Insurok) lightly sprung against the wall of the box, was exposed through this opening and controlled counting and reinforcing circuits.

The magazine was situated directly below the key and consisted of a metal door hinged at the top. This door was actuated by a cam on a constant-speed motor which completed a cycle in 5 sec. When a reinforcement was presented, the cam allowed the door to open abruptly, exposing the food and switching on a 7.5-w. lamp behind the food tray. The door remained open for 3.5 sec. and then closed gradually, turning off the light as it shut.

The external side of the key was illuminated by a 6-*v*., 18-amp. ribbon filament lamp whose beam was directed through a tunable wavelength filter and whose filament image was cast slightly out of focus on the key. The filter was a Cambridge Thermionic Corp. Monochromator (Model B). This instrument utilizes the differential rotation of quartz crystals for various wave lengths to produce dispersion of the spectrum. Its transmission spectrum at a given setting consists of a centroid of high relative intensity flanked symmetrically by a series of periodically spaced bands of lower intensity. Although the instrument may not be considered a true monochromator, it produces a stimulus patch of high apparent purity for the human observer. The band width of the centroid increases linearly from 14 to 24 Mμ over the range of settings from 450 to 640 Mμ. The percentage transmission of incident light is approximately flat over this range, varying between 2 and 4%. No attempt was made to produce an equal energy spectrum or alternatively to equate the spectrum lights in terms of the visibility function.

Procedure.—The procedure used in the present study derives from the work of Skinner (**11**) and Ferster (**3**), and is similar to that used by Brush, Bush, Jenkins, Johns, and Whiting (**2**). When the *Ss* were reduced to 80% of their ad lib. weight, they were trained to eat from the food magazine and were conditioned by the method of successive approximation to peck at the key. The 24 *Ss* were divided into groups of 6, and each group was trained on a given CS. The CS wave lengths were 530, 550, 580, and 600Mμ. Fifty consecutive reinforcements were given on each of two days. Following this, additional sessions of aperiodic reinforcement were administered, using a mean inter-reinforcement interval of approximately 1 min. During APR, 60-sec. stimulus-on intervals were alternated with 10-sec. stimulus-off intervals, and 30 stimulus-on intervals constituted each daily training session. During the stimulus-on interval a shutter in the light path of the monochromator was opened and the key was illuminated with colored light. At the same time, the interior of *S*'s box was illuminated by a 25-w. lamp reflected through the translucent plastic floor, resulting in an illuminance of <1.00 ft.-candle on the vertical walls. In the stimulus-off condition, similar to

TABLE 1

GENERALIZATION TEST STIMULI IN Mμ

Group	−70	−60	−50	−40	−30	−20	−10	CS	+10	+20	+30	+40	+50	+60
530		470		490	500	510	520	530	540	550	560	570		590
550		490		510	520	530	540	550	560	570	580	590		610
580		520		540	550	560	570	580	590	600	610	620		640
600	530		550	560	570	580	590	600	610	620	630	640		

the "blackout" situation used by Ferster (4), both the key and box illumination were terminated. The blackout condition was introduced to facilitate the changing of the key color during the stimulus generalization tests.

Generalization testing was carried out under extinction and was preceded by six 30-sec. intervals of responding to the CS during which three reinforcements were administered. The wave lengths used in the generalization tests are given in Table 1. The 11 different stimuli were randomized within a series and 12 different random series were presented to each *S* resulting in a schedule of 132 stimulus presentations. Twelve different schedules were constructed for *S*s given the 550 and 580 Mμ CS's and the same 12 schedules were subsequently used for *S*s given the 530 and 600 Mμ CS's. Each stimulus presentation was 30 sec. and was followed by a 10-sec. stimulus-off interval.

Following generalization testing, three additional APR sessions were given using the original CS. After this a second generalization test identical to the first was administered. During both testings, the number of key pecks for each 30-sec. test stimulus presentation was recorded.

RESULTS

Characteristics of the generalization gradient.—The mean generalization gradients obtained in the first test are shown in the upper portion of Fig. 1. For each CS group (530, 550, 580, and 600 Mμ), the gradient was obtained by plotting the mean total number of responses for each test stimulus against wave length in arithmetic units.

It will be seen that the various generalization gradients have highly comparable forms. On either side of the CS, the rate of responding declines first in a nearly linear fashion, and as the rate approaches zero, the curves become negatively accelerated. The differences among the curves for various CS's appear to be principally in terms of vertical displacement, i.e., total level of responding. The slopes of the generalization gradients, on the other hand, are very similar over the major portions of the curves and do not appear to conform to the expectation of marked changes corresponding to the transitions between spectral hues. The most conspicuous illustration of this point is the left half of the curve for 600 Mμ which passes from orange through yellow to green. It

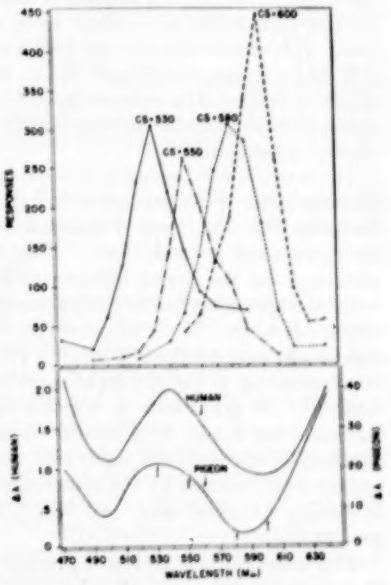

FIG. 1. Upper: Mean generalization gradients, first test. Lower: Hue discrimination as a function of wave length for pigeons (adapted from [5]), and humans (adapted from [1]).

exhibits approximately the same changes in slope as the other curves which do not pass through as many hues.

The similarity among the slopes of the curves is further substantiated by tests of trend (Type I analysis of variance [7]) for the right and left halves of the gradients separately. For the left half of all gradients, over the stimulus range from the CS to −40 Mμ, the interaction between wavelength of test stimuli and CS groups was not significant ($F = 1.02$; 12, 80 *df*), indicating that the hypothesis of parallel slope cannot be rejected over this range. For the right side of these gradients, over the stimulus range from the CS to +40 Mμ, the interaction was significant beyond the .1% level ($F = 3.84$; 12, 80 *df*). A further analysis of the differences in slope between pairs of gradients on the right side revealed that the interaction previously obtained resulted from a slope difference between the 600 and 580 Mμ CS groups ($F = 5.10$; 4, 40 *df*; $P < .005$). The other slope differences on the right side were not statistically significant.

Evidence for differences in vertical displacement of the gradients was also furnished by the Type I analysis of variance used to test slope. The *F* obtained for the mean differences in rate of responding for the left sides of the gradient is 5.55 (3, 20 *df*) which is significant beyond the 1% level. The corresponding *F* for the right sides is 2.80 (3, 20 *df*) which is significant between the 5 and 10% levels. The mean differences on the right side are presumably reduced by the interaction involving the 600 and 580 Mμ CS groups.

The lower portion of Fig. 1 presents the spectral difference threshold ($\Delta\lambda = f(\lambda)$) for both pigeons (**S**) and humans (**1**) adjusted to the same approximate ordinates. The functions for the two species appear to be highly similar with minima and maxima at approximately the same wave lengths. Three characteristics of the $\Delta\lambda$ function might be reflected in the generalization gradients. The first is the absolute value of $\Delta\lambda$ which determines the jnd interval between any two values on the continuum, and hence might be related to the slope of the gradient. The other two characteristics are the first and second derivatives of $\Delta\lambda$ with respect to λ, which might be related to rate of change of curvature. For the purpose of the present analysis, we shall be concerned mainly with the absolute value of $\Delta\lambda$.

In terms of the discriminability function for the pigeon, the generalization gradients for CS values of 530 and 580 Mμ should be relatively symmetrical. For values near the CS, however, the gradient for 580 Mλ should be steeper than for any other CS. The gradients centered at 550 and 600 Mμ should be asymmetrical, with 550 Mμ being steeper in the direction of increasing wave length, and vice versa in the case of 600 Mμ.

In general, these expectations are not borne out. Figure 1 and the preceding statistical analyses of slope differences suggest that for the most part the gradients are of uniform slope. The major exceptions to uniformity of slope are the relative steepness of the 600 Mμ gradient and the flatness of the 580 Mμ gradient, both of which features are in the direction contrary to the hypothesis based on the $\Delta\lambda$ function.

The bidirectional symmetry of the gradients was examined by means of separate analyses of variance (A × B × S [7]) for each CS group. In effect, one side of the gradient was superimposed on the other and a test of trend difference was obtained from the A × B (Right-Left Sides

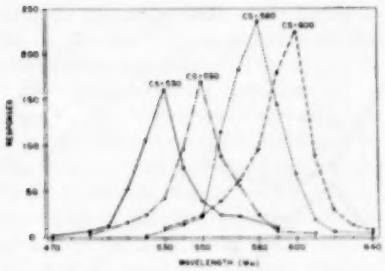

FIG. 2. Mean generalization gradients, second test.

× Test Stimuli) interaction. The values entering into this analysis were obtained by subtracting response rate for each test stimulus from the response rate at the CS. The hypothesis of symmetry could not be rejected for the 530 and 580 gradients, since the interaction F in each case was less than 1.00. This appears to conform to the $\Delta\lambda$ hypothesis. On the other hand, however, significant asymmetries contrary to the direction indicated by the $\Delta\lambda$ function are revealed in the analyses of variance for the 550 and 600 Mμ groups. The interaction F for the 550 Mμ gradient is 8.39, which is significant beyond the .1% level for 4, 20 *df*; the F for the 600 Mμ gradient is 5.85, which is significant beyond the 1% level for 3, 15 *df*.[3]

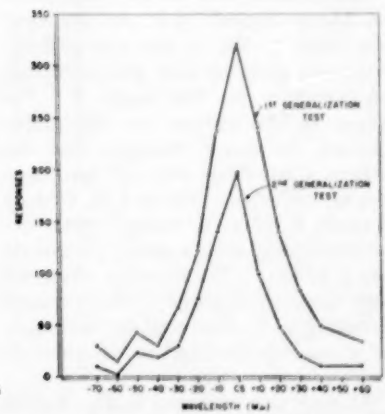

FIG. 3. Mean generalization gradients, first and second test, for all CS groups combined.

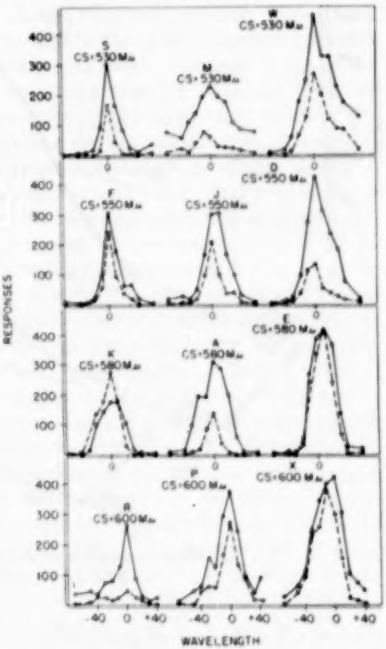

FIG. 4. Individual generalization gradients for 12 *S*s representing various CS groups and levels of response strength. Solid lines designate the first test, broken lines the second test.

The results of the second generalization test, shown in Fig. 2, are highly similar to those obtained in the first test. In general, all gradients are reduced in height and the differences in vertical displacement and slope are attenuated. Neither the F for vertical displacement nor the F for slope difference is significant ($P \gtrsim .20$). The hypothesis of symmetry can be rejected only for the 600 Mμ gradient, since $F = 3.44$ (3, 15 *df*; $.025 < P < .05$). The relationships among the

[3] The number of *df* is reduced for the 600 Mμ group because only four points to the right of the CS were available for analysis.

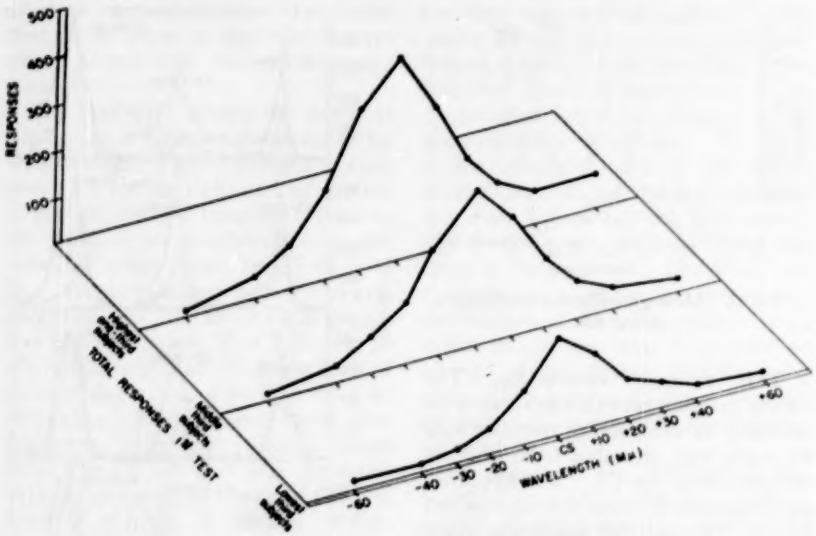

FIG. 5. Mean generalization gradients for groups of eight *S*s differing in total number of responses, first test.

gradients for the second generalization test are such that the curves can be nearly superimposed by translating them along the wave-length axis to a common CS point.

The relationships between the mean generalization gradient for all groups in the first test and the mean for all groups in the second test are shown in Fig. 3.

Individual generalization gradients.—Twelve pairs of individual generalization gradients for the first and second tests are shown in Fig. 4. These gradients are arranged in order of ascending response rate, and for each CS group the *S*s presented are the lowest, the highest, and the *S* nearest the group median. Perhaps the most salient aspect of these curves is the orderliness and reproducibility of the generalization process within the individual *S*. Although these gradients are summed over the individual's schedule of 132 stimulus presentations, it is interesting that a similar bidirectional gradient may be obtained from any one series of 11 stimulus presentations. A second interesting feature of these curves is the extent to which each *S* produces a gradient in the second test which is almost a replica of the first.

An examination of the individual gradients suggests that the averaged gradients in Fig. 3 are not entirely representative of the generalization phenomenon for the single *S*. For some *S*s the curves are bilaterally convex, for some, concave, and for others, concave on one side and convex on the other. Certain *S*s, such as Animal E (Fig. 4, second test), exhibit linearity over a major portion of the gradient. The linearity observed over the central range of the averaged curve (Fig. 3) may well be the result of a random distribution of concavities and convexities.

It may also be noted in Fig. 4 that certain *S*s show maximum responding at a point either ± 10 Mμ from the CS.

In some instances this contributes to the departure from symmetry of the averaged group curves (Fig. 1).

The generalization gradient as a function of differences in response strength.—The averaged gradients in Fig. 5 were obtained by dividing the 24 *Ss* (without respect to CS group) into three relatively homogeneous subgroups in terms of total responses for the first generalization test. As the total response level is reduced, the generalization gradient becomes uniformly flatter. Tests of trend differences over the right and left halves of the gradients separately permit rejection of the hypothesis that the curves are parallel. On the right side, the interaction $F = 22.4$ (8, 84 *df*; $P < .001$) while F for vertical displacement is 19.8 (2, 21 *df*; $P < .001$). The interaction F for the left halves is 5.15 (8, 84 *df*; $P < .001$), while the vertical displacement $F = 19.5$ (2, 21 *df*; $P < .001$). In a later section we shall analyze in greater detail the nature of these interactions.

Additional information concerning changes in the generalization gradient associated with changes in level of responding may be obtained from an analysis of the extinction process occurring during the generalization test. The generalization gradients in Fig. 6 are based upon the total responses of all *Ss* and represent the average performance for successive blocks of 33 stimulus presentations. The gradient for the first quarter of the extinction series is relatively steep, and the gradients for the succeeding quarters become progressively flatter. These changes bear a strong resemblance to those shown by the response-strength subgroups in Fig. 5. Furthermore, the gradients for the successive stages of extinction (Fig. 6) display a departure from parallelism on both sides of the CS. The interaction F's for the right and left halves of the gradients are 8.22 and 6.40, both significant beyond the .1% level for 12, 276 *df*.

The similarities between the changes in the form of the generalization

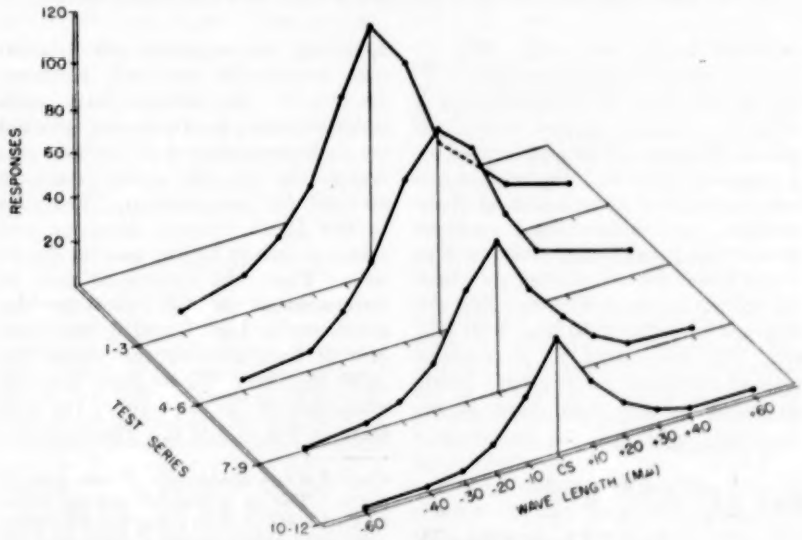

FIG. 6. Mean generalization gradients for successive fourths of first test.

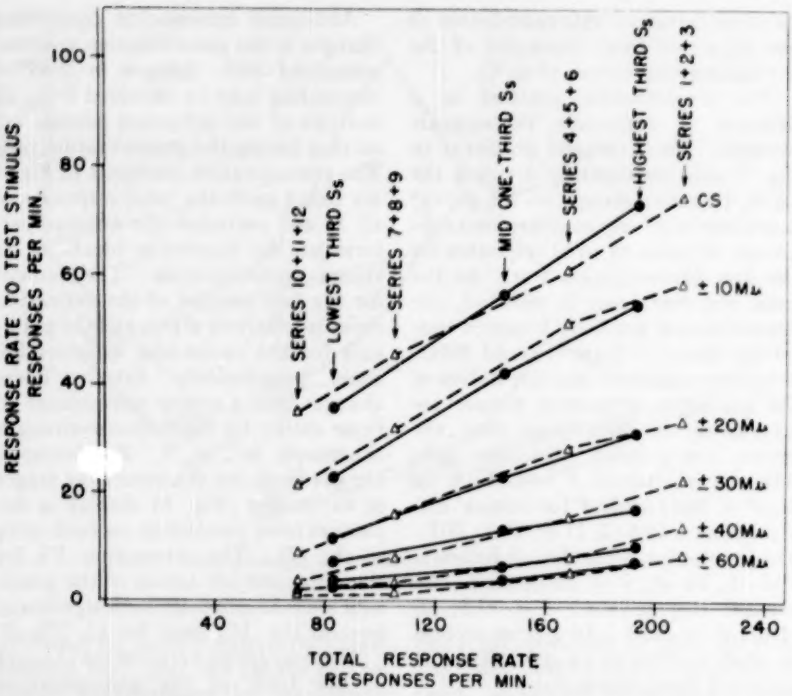

FIG. 7. Rate of responding for various test stimuli as a function of total response strength.

gradient during extinction (Fig. 6) and the changes associated with individual differences in response strength (Fig. 5) strongly suggest that both sets of changes may be represented by a common function. In order to obtain a graphical comparison of these changes, each bidirectional gradient for the first generalization test in Fig. 5 and 6 was first converted to a unidirectional gradient by averaging the response rates for ± 10 Mμ, ± 20 Mμ, etc., and converting to a common scale of responses per minute. Next, an index[4] of the total rate of responding (in responses per minute) was determined for each gradient. In Fig. 7 the values from each unidirectional gradient are plotted on ordinates erected at points corresponding to the total response strength for that gradient. The lines in this figure connect the same test stimulus values for the various gradients. Thus, the uppermost pair of lines connect the CS values for the gradients in Fig. 5 and 6, the next pair of lines connect the values for ± 10 Mμ, etc. These lines may be visualized as the contours of the surfaces in Fig. 5 and 6. The high de-

[4] The index used was

$$R_{CS} + \sum_{n=1}^{6} \left(\frac{[R_{+10M\mu} + R_{-10M\mu}]}{2} + \ldots \frac{[R_{+60M\mu} + R_{-60M\mu}]}{2} \right)$$

where R = rate of responding in responses per minute. This index provides that the values for the CS will lie along a line with a 0,0 intercept, and permits a rescaling of the total extinction curve on a response strength axis.

gree of correspondence between these lines indicates that changes in generalization as a function of response strength are the same whether the response strength varies within the individual, as in the course of extinction, or whether response strength is considered as a parameter characterizing the differences between individuals.

The transformation of the data to the form shown in Fig. 7 suggests the mathematical operation necessary to predict generalization at one level of response strength given the generalization function at a different level. The operation proposed by Hull (**6**) is multiplication by a constant. Miller (**9**) has suggested, in the context of a discussion of the effect of motivational changes on the gradient, that the conversion can be accomplished by adding a constant at every point on the gradient. If the multiplicative relation holds, then all the lines in Fig. 7 should pass through the point 0, 0, but if the additive relation obtains, then all the lines in this figure should be parallel to each other. The present results indicate that neither hypothesis alone provides a sufficient description of the situation. The additive relation appears to hold for values near the CS, but the multiplicative relation must be invoked for values further removed from the CS.

Discussion

The results of the present experiment do not appear to provide evidence for a correspondence between the discriminability function for wave length of light ($\Delta\lambda$ vs. λ) and the generalization process. Since some relationship is reasonable on logical and intuitive grounds, our results must be interpreted in the light of a number of possibilities. In the first place, the relationship may not, in fact, exist, because during generalization testing the *S* is not making a comparative judgment. In more formal terms, the analogy we have suggested between the generalization decrement and the difference limen may be misleading, because $p(R)$ in one situation refers to a simple conditioned response, while in the psychophysical experiment, $p(R)$ is associated with a complex of discriminatory responses.

The foregoing considerations, however, fall far short of excluding the possibility that a relationship does exist and that the present experiment has failed to disclose it. Our results are compatible with the notion that the wave-length discriminability function for the pigeon is much flatter than has been reported (**5**), flatter in fact than that for humans. It may also be that the discriminability differences in our situation are attenuated because the pigeon works in a dim light. As a check on the latter possibility a crude psychophysical experiment was performed on four human *S*s using the method of minimal changes. The $\Delta\lambda$ function obtained was not appreciably different from that generally reported; it showed minima in the blue and yellow regions, a maximum in the green, and differences on the order of 5:1. The same results were obtained in the dark and in the presence of dim light.

Another factor in our experiment which might act to distort the expected relationship is brightness variation, since no attempt was made to equate the test stimuli for brightness or incident energy. This might have the consequence of increasing the slope of gradients passing into regions of lower visibility, and might account for the asymmetry of the gradient for 600 Mμ. If, however, brightness were an important factor over all, it would be difficult to account for the high degree of similarity among the gradients plotted on a wave-length continuum. The extent to which brightness can be invoked as a variable is inversely proportional to the degree of relationship between changes in response rate and wave length.

Finally, serious consideration must be given to the possibility that the relationship between generalization and discrimination is much more complex than

we have supposed. Before any general statement can be made concerning the independence of discriminability and stimulus generalization, the problems raised by the present experiment must be explored with other continua, species, and techniques.

With respect to the issue of changes in the form of the generalization gradient with changes in response strength, our findings may be compared most directly with those of Margolius (**8**), who investigated changes in generalization as a function of number of conditioning trials. In terms of the rate of responding measure, Margolius found increases in the absolute and relative amounts of generalization as the number of training trials increased. Margolius' interpretation of his findings is equivalent to what we have designated as the additive relation. It will be recalled that for the present results, the additive hypothesis was tenable for the major area of the curve including values near the CS, and high response strengths, but for remote values of the stimuli and for low response strengths the additive relation fails. This suggests the possibility that the empirical gradient is comprised of two components, one which is linear on a wave-length scale and changes additively, and another which is curvilinear and changes multiplicatively. It may be that the second component is contributed by random responding and responding to extraneous stimuli accompanying the beginning of a trial. This separation of components implies that if the contribution of the second component was reduced, either experimentally or analytically, the additive relation would become more conspicuous.

Summary

An experiment was performed to test the hypothesis of a relationship between the discriminability of spectrum colors and stimulus generalization along the wave-length continuum. Four groups of six pigeons were trained to peck at a disc illuminated by monochromatic lights corresponding to 530, 550, 580, and 600 Mμ. Bidirectional generalization gradients were obtained from measures of response rate during extinction. These gradients were of highly comparable forms for the various spectral regions tested. The results open the possibility of an independence between the generalization decrement and the discriminability of stimuli.

Changes in the generalization gradient accompanying changes in response strength were investigated. A close correspondence was found between changes in the form of the gradient during extinction and changes associated with individual differences in response strength. In both cases the changes appear to be describable by means of a combined additive and multiplicative relationship.

References

1. Boring, E. G., Langfeld, H. S., & Weld, H. P. *Foundations of psychology.* New York: Wiley, 1948.
2. Brush, F. R., Bush, R. R., Jenkins, W. O., John, W. F., & Whiting, J. W. M. Stimulus generalization after extinction and punishment: an experimental study of displacement. *J. abnorm. soc. Psychol.*, 1952, **47**, 633–640.
3. Ferster, C. B. The use of the free operant in the analysis of behavior. *Psychol. Bull.*, 1953, **50**, 263–274.
4. Ferster, C. B. Sustained behavior under delayed reinforcement. *J. exp. Psychol.*, 1953, **45**, 218–224.
5. Hamilton, W. F., & Coleman, T. B. Trichromatic vision in the pigeon as illustrated by the spectral hue discrimination curve. *J. comp. Psychol.*, 1933, **15**, 183–191.
6. Hull, C. L., *Principles of behavior.* New York: D. Appleton-Century, 1943.
7. Lindquist, E. F. *Design and analysis of experiments in psychology and education.* New York: Houghton Mifflin, 1953.
8. Margolius, G. Stimulus generalization of an instrumental response as a function of the number of reinforced trials. *J. exp. Psychol.*, 1955, **49**, 105–111.
9. Miller, N. E. Experimental studies of conflict. In J. McV. Hunt (Ed.), *Personality and the behavior disorders.* New York: Ronald, 1944.
10. Schlosberg, H., & Solomon, R. L. Latency of response in a choice discrimination. *J. exp. Psychol.*, 1943, **33**, 22–39.
11. Skinner, B. F. Are theories of learning necessary? *Psychol. Rev.*, 1950, **57**, 193–216.
12. Spence, K. W. The differential response in animals to stimuli varying within a single dimension. *Psychol. Rev.*, 1937, **44**, 430–444.

(Received for early publication August 8, 1955)

Psycholinguistics

A Survey of Theory and Research Problems

Report of the 1953 Summer Seminar Sponsored by the Committee on Linguistics and Psychology of the Social Sciences Research Council

CHARLES E. OSGOOD, *Editor* THOMAS A. SEBEOK, *Associate Editor*

With a Foreword by JOHN W. GARDNER

JOHN B. CARROLL
SUSAN M. ERVIN
JOSEPH H. GREENBERG
JAMES L. JENKINS
FLOYD G. LOUNSBURY
LEONARD D. NEWMARK
SOL SAPORTA
DONALD E. WALKER
KELLOGG WILSON

Price, $2.50

Order from

AMERICAN PSYCHOLOGICAL ASSOCIATION
1333 Sixteenth Street N. W.
Washington 6, D. C.

QUARTERLY JOURNAL OF EXPERIMENTAL PSYCHOLOGY

Founded in 1948

Published Quarterly in February, May, August, and November

for the

EXPERIMENTAL PSYCHOLOGY GROUP

Editor: D. Russell Davis, University of Cambridge

Subscriptions from the United States
should be entered now for the next year.
30 shillings per year (current rate $4.20).

Order from

W. Heffer & Sons, Ltd.
Cambridge, England